

E.T.S. de Ingeniería Industrial,
Informática y de Telecomunicación

Aplicación de la inteligencia de negocio mediante la herramienta Pentaho



Grado en Ingeniería Informática

Trabajo Fin de Grado

Autor: Ramiro Marín Sesma

Tutor: José Javier Astrain Escola

Cotutor: Alberto Córdoba Izaguirre

Pamplona, 29/06/2015



Agradecimientos

Quiero aprovechar esta ocasión para agradecer a todas las personas que han hecho posible que yo llegue hasta este punto en mi vida. No ha sido fácil, pero nadie dijo que lo fuera a ser. Por el camino he sufrido cantidad de baches y contratiempos, los cuales, uniendo los puntos hacia atrás en el tiempo, veo que han sido más que beneficiosos para mi desarrollo como persona y me alegro de haberlos sufrido y superado. “El que tropieza y no cae, da un pequeño salto y adelanta”.

En primer lugar agradecer a mis tutores Alberto y José Javier la confianza depositada en mí a la hora de hacer este trabajo y su predisposición a la hora de solventar mis dudas, así como su orientación formativa y laboral.

También quiero agradecer profundamente el esfuerzo económico y psicológico que han hecho mis padres, Julia y Ramiro, para ofrecerme la posibilidad de tener unos estudios. Por supuesto a mis hermanos Javier y Álvaro les agradezco la confianza que tienen en mí y lo que me valoran como persona, estoy seguro de que ellos llegarán a ser mucho mejor que yo en todo.

Gracias a mis abuelos Ramón y Rosario, piezas clave de esta familia, por enseñarme desde la sabiduría que les aportan los años a ver las cosas desde una perspectiva adecuada. “Más sabe el diablo por viejo que por diablo”.

Gracias a mis tíos Arnaldo y Miriam por darme tanto cariño y comprensión y aconsejarme siempre en mis peores momentos ayudándome a tomar las decisiones más adecuadas.

Gracias a mi tío Álvaro por ofrecerme siempre una ayuda a la hora de canalizar mis inquietudes.

Gracias a mi tío Pedro Joaquín por estar siempre ahí, exigiéndome pero también animándome cuando es necesario, por ofrecerme una visión crítica del mundo en el que vivimos. Gracias tío también por ayudarme a desarrollar mis conocimientos en otras ramas, como es la Economía. “Un hombre solo envejece cuando deja de aprender”

Gracias a mi tía Emilita por animarme siempre que tengo la ocasión de hablar con ella, sé que aunque no compartamos mucho tiempo también me apoya y me anima en todo. Quiero dar las gracias también a mi tío Pedro José, que aunque ya no se encuentre entre nosotros, su alegría, forma de ver la vida y conocimientos han llegado a mí a través de mi padre.

Gracias a toda la familia Mena por apoyarme siempre en los malos momentos y en especial a Inma por ayudarme a completar satisfactoriamente mis estudios, por su paciencia a la hora de enseñarme y por supuesto por su cariño.

Agradecer, por último y no por ello menos importante, a mis buenos amigos sus buenos consejos en todo lo relativo a mi orientación formativa, además del apoyo prestado en tantas otras adversidades que me han ayudado a superar haciéndome mejor persona.

En general a todos vosotros os quiero agradecer la confianza que habéis depositado en mí, la mayoría de las veces mayor que la mía propia.

Palabras clave y definiciones previas

- **Dato:** Es una representación simbólica de una variable cuantitativa. Un dato en sí mismo no ofrece ningún tipo de información. Ejemplo: "15".
- **Información:** Cuando es posible obtener un conjunto organizado de los datos y así obtener algún tipo de conocimiento acerca de ellos podemos hablar de información. Ejemplo: "Hay 15 lince ibéricos."
- **Conocimiento:** Es posible transformar la información en conocimiento cuando se dispone de unos datos previos sobre el mismo suceso, experiencia. Ejemplo: "Hay 15 lince ibéricos, por lo que la población ha disminuido respecto al año pasado"
- **Business Intelligence:** Business Intelligence, o Inteligencia de Negocio en español, engloba todo el conjunto de técnicas necesarias para poder transformar los datos e informaciones de una empresa en conocimiento con el cual poder tomar decisiones estratégicas, por ejemplo utilizando ese conocimiento en un análisis DAFO (Debilidades, Amenazas, Fortalezas y Oportunidades).
- **Proceso ETL (Extract, Transform and Load):** Es el proceso mediante el cual se extraen datos desde cualquier fuente de datos relevante para la empresa, se transforman como sea necesario y posteriormente se cargan en otra base de datos o Data Warehouse para realizar el posterior proceso de análisis.
- **Cubo OLAP:** Son las siglas en inglés de *OnLine Analytical Processing*, es decir, procesamiento analítico en línea. Conceptualmente es una base de datos multidimensional en la cual el almacenamiento físico de los datos se realiza también en un vector multidimensional. Es decir, tendremos una tabla de hechos, en donde aparecerá la información sobre la que queramos medir en cualquiera de sus dimensiones y enlazadas a esta, por medio de clave foránea aparecerán las dimensiones de medida del cubo, que pueden ser tantas como queramos (>3 dimensiones Hipercubo), aunque cabe destacar que teniendo más de tres dimensiones se hace difícil de comprender espacialmente, puesto que la mente humana está acostumbrada a trabajar con únicamente tres dimensiones. Fue propuesto por Edgar Frank Codd para solucionar el lento análisis sobre bases de datos relacionales más optimizadas para el registro de transacciones u OLTP (OnLine Transaction Processing).

- **Data Warehouse:** Un Data Warehouse o almacén de datos es un conjunto de datos de un determinado contexto, integrado, no volátil y variable en el tiempo. Principalmente se trata de un histórico de datos de la empresa, más allá de la información operacional o transaccional. Es una base de datos orientada al análisis. Ejemplo: "Histórico de ventas de una empresa en los últimos 25 años".
- **Reporte:** Un reporte es un informe que organiza y exhibe la información que se encuentra contenida en una base de datos. El reporte pretende dotar de mayor utilidad a los datos, además de hacerlo más interpretable para el usuario.

Índice de contenido

1	Introducción	10
1.1	Introducción	10
1.2	Business Intelligence (Inteligencia de Negocio).....	11
1.3	Contexto tecnológico	13
1.4	Metodología empleada	14
2	Instalación y configuración de Pentaho Community.	15
2.1	Descarga y configuración de componentes	15
2.1.1	Java Development Kit.....	15
2.1.2	Configuración de las variables de entorno	15
2.1.3	MySQL.....	18
2.1.4	Pentaho BI Server.....	18
2.1.5	Data Sources	19
2.2	Resto de componentes de Pentaho.....	23
2.3	Instalación del Marketplace de Pentaho	24
2.4	Instalación de Saiku Analytics 2.6 (DEPRECIATED).....	24
2.5	Conclusión	25
3	Proceso ETL y creación del cubo OLAP.	26
3.1	Proceso ETL	26
3.1.1	Nueva transformación.....	27
3.1.2	Entrada Excel.....	28
3.1.3	Salida tabla	30
3.1.4	Vista general del ETL	33
3.2	Creación de cubo OLAP con Schema Workbench	35
3.3	Visualización del cubo OLAP con jPivot Wiew	41
4	Construcción del Dashboard.....	42
4.1	Consideraciones para construir un gráfico adecuado.....	42
4.2	Creación de Dashboard	43
4.2.1	Layout Structure.....	44
4.2.2	Datasources	45
4.2.3	Components.....	51

4.3	Vista del Dashboard	53
4.4	Utilización del Dashboard.....	54
5	Elaboración de Reportes.....	56
5.1	Realización del reporte.....	56
5.1.1	Look and Feel	57
5.1.2	Data Source and Query	58
5.1.3	Layout Step	63
5.2	Conclusión	66
6	Conclusiones y líneas futuras.....	67
7	Referencias	68

Listado de Ilustraciones

<i>Ilustración 1: Inicio Equipo (clic derecho), Propiedades.....</i>	<i>15</i>
<i>Ilustración 2: Panel de control, Configuración avanzada del sistema</i>	<i>16</i>
<i>Ilustración 3: Variables de entorno</i>	<i>16</i>
<i>Ilustración 4: Path</i>	<i>17</i>
<i>Ilustración 5: PENTAHO_JAVA_HOME</i>	<i>17</i>
<i>Ilustración 6: Pantalla inicial Pentaho.....</i>	<i>19</i>
<i>Ilustración 7: Manage Data Sources</i>	<i>19</i>
<i>Ilustración 8: New Data Source 1</i>	<i>20</i>
<i>Ilustración 9: New Data Source 2</i>	<i>20</i>
<i>Ilustración 10: Credenciales MySQL</i>	<i>21</i>
<i>Ilustración 11: Test Data Source</i>	<i>21</i>
<i>Ilustración 12 Data Source OK 1.....</i>	<i>22</i>
<i>Ilustración 13: Data Source OK 2.....</i>	<i>23</i>
<i>Ilustración 14: Marketplace</i>	<i>24</i>
<i>Ilustración 15: Saiku Analytics.....</i>	<i>24</i>
<i>Ilustración 16: Saiku Analytics Versión 2.6</i>	<i>25</i>
<i>Ilustración 17: Pantalla inicial Pentaho Data Integration</i>	<i>26</i>
<i>Ilustración 18: Fichero, Nuevo, Transformación</i>	<i>27</i>
<i>Ilustración 19: Entrada Excel 1</i>	<i>28</i>
<i>Ilustración 20: Entrada Excel 2</i>	<i>29</i>
<i>Ilustración 21: Entrada Excel 3</i>	<i>29</i>
<i>Ilustración 22: Entrada Excel 4</i>	<i>30</i>
<i>Ilustración 23: Salida Tabla</i>	<i>30</i>
<i>Ilustración 24: Database Connection ETL.....</i>	<i>31</i>
<i>Ilustración 25: Database Connection Test ETL</i>	<i>31</i>
<i>Ilustración 26: Salida Tabla 2</i>	<i>32</i>
<i>Ilustración 27: Vista General ETL.....</i>	<i>33</i>
<i>Ilustración 28: Ejecutar transformación.....</i>	<i>34</i>
<i>Ilustración 29: Topología en estrella.....</i>	<i>34</i>
<i>Ilustración 30: Database Connection Schema Workbench</i>	<i>35</i>
<i>Ilustración 31: File, New, Schema</i>	<i>35</i>
<i>Ilustración 32: Add Cube</i>	<i>36</i>
<i>Ilustración 33: Add Dimensions.....</i>	<i>36</i>
<i>Ilustración 34: Dimensions</i>	<i>37</i>
<i>Ilustración 35: Add Table</i>	<i>37</i>
<i>Ilustración 36: Table.....</i>	<i>38</i>
<i>Ilustración 37: Add Level</i>	<i>38</i>
<i>Ilustración 38: Level</i>	<i>39</i>
<i>Ilustración 39: Measures.....</i>	<i>39</i>

<i>Ilustración 40: File, Publish</i>	40
<i>Ilustración 41: Crear nuevo, JPivot View</i>	41
<i>Ilustración 42: New JPivot View</i>	41
<i>Ilustración 43: JPivot View</i>	41
<i>Ilustración 44: Comparativa de gráficos</i>	42
<i>Ilustración 45: Pantalla inicial Pentaho BI Server</i>	43
<i>Ilustración 46: Create New, CDE Dashboard</i>	44
<i>Ilustración 47: Layout Structure</i>	44
<i>Ilustración 48: Datasources</i>	45
<i>Ilustración 49: OLAP Selector Wizard 1</i>	45
<i>Ilustración 50: OLAP Selector Wizard 2</i>	46
<i>Ilustración 51: Saiku OLAP Wizard 1</i>	47
<i>Ilustración 52: Saiku OLAP Wizard 2</i>	48
<i>Ilustración 53: Saiku OLAP Wizard 3</i>	48
<i>Ilustración 54: Jndi y Mondrian Schema</i>	49
<i>Ilustración 55: Query y Parameters</i>	49
<i>Ilustración 56: Query</i>	50
<i>Ilustración 57: Parameters</i>	50
<i>Ilustración 58: Charts</i>	51
<i>Ilustración 59: Modificaciones Chart</i>	52
<i>Ilustración 60: Listeners Chart</i>	52
<i>Ilustración 61: Parameters Chart</i>	52
<i>Ilustración 62: Vista rápida del Dashboard</i>	53
<i>Ilustración 63: Utilización del Dashboard</i>	54
<i>Ilustración 64: Pantalla inicial Pentaho Report Designer</i>	56
<i>Ilustración 65: Report Design Wizard Look and Feel</i>	57
<i>Ilustración 66: Report Design Wizard Data Source and Query 1</i>	58
<i>Ilustración 67: Report Design Wizard Data Source and Query 2</i>	58
<i>Ilustración 68: Query Database Connection</i>	59
<i>Ilustración 69: Query Database Connection Test</i>	60
<i>Ilustración 70: New Query 1</i>	60
<i>Ilustración 71: SQL Query Designer</i>	61
<i>Ilustración 72: New Query 2</i>	62
<i>Ilustración 73: Report Design Wizard Data Source and Query 3</i>	63
<i>Ilustración 74: Report Design Wizard Layout Step</i>	63
<i>Ilustración 75: Vista general Reporte</i>	64
<i>Ilustración 76: Run Report</i>	65
<i>Ilustración 77: Print Preview Report</i>	65

1 Introducción

1.1 Introducción

En esta memoria se presenta la aplicación de un proyecto completo de Business Intelligence, desde el proceso ETL hasta la creación de un Dashboard dinámico e intuitivo que permita realizar un análisis ágil con el objetivo de tomar una decisión.

Para ello se ha utilizado la herramienta Pentaho Community que, además de ser gratuita, aporta una funcionalidad más que considerable para realizar todo el proceso planteado. Como base de datos para nutrir a Pentaho se ha elegido MySQL Server puesto que es una herramienta que aporta la funcionalidad requerida y con la cual ya se estaba familiarizado.

En este caso se han elegido los datos de comercio exterior disponibles en la página web del INE [1] (Instituto Nacional de Estadística) para realizar el proyecto pero cabe destacar, que el proceso de Business Intelligence es independiente del contexto de los datos que se utilicen, se puede realizar para datos de cualquier temática. Siempre con la intención de ofrecer una herramienta que permita realizar un análisis de las distintas dimensiones de medida del cubo OLAP sobre la tabla de hechos en tiempo real y de la manera más ágil e intuitiva posible.

Se debe tener en cuenta que el campo más utilizado de estas aplicaciones es la gerencia de empresas en dónde la decisión de un directivo puede cambiar el rumbo completo de la compañía produciendo así un beneficio para ésta o por el contrario un decremento del mismo a causa de una mala decisión. Si bien es cierto que no toda la responsabilidad de decisión recae sobre el Sistema de Business Intelligence, éste debe ser lo más preciso y adecuado posible para favorecer una resolución acertada.

1.2 Business Intelligence (Inteligencia de Negocio)

Business Intelligence, o Inteligencia de Negocio en español, engloba todo el conjunto de técnicas necesarias para poder transformar los datos e informaciones de una empresa, o cualquier otro supuesto del que se tengan evidencias, en conocimiento con el cual apoyar la toma de decisiones de una manera ágil, correcta e intuitiva.

En concreto, en el mundo empresarial los datos se extraen de la producción de la compañía, analizando los aquellos a partir de los que se puedan tomar decisiones estratégicas de futuro con el objetivo de que la compañía mejore sus beneficios y su competitividad frente al resto de empresas del sector.

Algunas de las distintas herramientas de Business Intelligence de las que se disponen actualmente en el mercado son las siguientes:

Software libre:

- Pentaho CE [2].
 - Desarrollo en Java.
 - La más extendida y utilizada.
 - Dispone de cubos OLAP.
 - Multiplataforma.
- RapidMiner [3].
 - Desarrollo en Java.
 - Multiplataforma.
 - Dispone de una API de Java.
- SpagoBI [4].
- Eclipse BIRT Project [5]: Generador de informes para aplicaciones Web de código abierto basado en Eclipse.
- OpenI [6]: Aplicación Web simple orientada al reporting OLAP.
- Palo [7].
 - El entorno de trabajo es Excell
 - Permite intercambiar datos con SAP

Software comercial:

- Informe SQL, QLR Manager en español [8].
- Oracle BI [9].
- Microsoft SQL Server BI Edition [10]: Analysis Services, Integration Services y reporting Services.
 - Alta compatibilidad con productos Microsoft.
- Sap Business Objects [11].
- Panorama Software [12].

- Pentaho BI_SW [13].
 - Dispone de cubos OLAP.
 - Visualización desde iOS y Android.

Según el grado de complejidad, la inteligencia de negocio se puede clasificar en distintos niveles:

- Informes:
 - Informes predefinidos: Se trata de informes fijados en los que se muestran siempre los mismos datos. Es necesario una persona especializada que los realice.
 - Informes a medida: Se trata de informes en los que el usuario define los datos que desea mostrar en el momento en el que se realiza el informe.
 - Consultas (Query) / Cubos OLAP (On-Line Analytic Processing): Se trata de consultas en tiempo real realizadas sobre los cubos anteriormente establecidos por un técnico.
- Análisis:
 - Análisis estadístico: Es capaz de ofrecer con los datos seleccionados análisis estadísticos para cumplimentar en profundidad la información.
 - Pronósticos (Forecasting): Se trata de ofrecer conocimientos anticipados de algún suceso, como es lógico todo pronóstico está ligado a un margen de error en su predicción.
 - Modelado predictivo o Minería de datos (Data Mining): La minería de datos intenta descubrir patrones o comportamientos en una gran cantidad de datos con el objetivo de definir un comportamiento común.
 - Optimización: Ofrecen soluciones de optimización sobre los datos analizados.
 - Minería de procesos: Se centra en analizar los procesos de negocio de acuerdo con un registro de eventos para dar con comportamientos de negocio definidos que se puedan mejorar.

Para concluir, se podría decir que en general todas las herramientas de Business Intelligence deben tener las siguientes características:

- Información accesible. Acceso a los datos con independencia de su procedencia.
- Apoyar la toma de decisiones. Con BI se busca ir más allá de la mera presentación de la información, se le debe proporcionar al usuario herramientas de análisis selectivo de datos.
- Orientadas al usuario final. Con BI se busca independencia entre la parte técnica necesaria para la creación de un proyecto de BI y el usuario final que utilice dicha aplicación que no necesariamente debe disponer de conocimientos técnicos para poder utilizarla adecuadamente.

1.3 Contexto tecnológico

Se ha elegido la herramienta Pentaho Community puesto que al ser una herramienta open source es posible trabajar con ella sin ningún tipo de licencia de pago. Además de eso ofrece una funcionalidad suficientemente potente como para realizar todo lo necesario para este proyecto. Dentro de las herramientas de software libre disponibles en el mercado Pentaho Community es la más extendida y utilizada lo que aporta una ventaja a la hora de documentarse y aprender sobre la misma.

Cabe destacar que Pentaho es una herramienta de Business Intelligence desarrollada en Java, y ejecutada en un servidor Tomcat, lo cual hace muy accesible la aplicación puesto que es posible utilizarla desde cualquier dispositivo y desde cualquier sistema operativo. Motivo realmente importante cuando se dispone de empleados que utilizan distintos sistemas operativos y distintos dispositivos, esto permite que todos puedan acceder indistintamente a Pentaho y realizar las labores pertinentes de una manera adecuada y sin ningún tipo de incompatibilidad.

Pentaho Community es un conjunto de programas de software libre que nos permiten realizar todo el proceso de Business Intelligence a partir de un conjunto de información proveniente de cualquier tipo de almacenamiento estructurado, como bien puede ser una base de datos, una hoja de cálculo, un fichero de texto, un archivo JSON (*JavaScript Object Notation*)... En definitiva, casi cualquier tipo estructurado de datos servirá para realizar un proyecto de Business Intelligence con esta herramienta. Cabe destacar que Pentaho Community está compuesto por varios módulos o herramientas:

- *Business Analytics Platform: Plataforma para el análisis.
- *Data Integration: Herramienta de Pentaho para ETL (Extraction Transformation and Load).
- *Report Designer: Herramienta destinada a la creación y diseño de informes.
- Marketplace: Market de aplicaciones de terceros y propias de Pentaho que pueden aportar una funcionalidad extra.
- Aggregation Designer: Creación y manejo de tablas agregadas.
- *Schema Workbench: Con esta herramienta podremos realizar los cubos MOLAP (*Multidimensional OnLine Analytical Processing*).
- Metadata Editor: Herramienta que permite crear un modelo derivado de un Data Warehouse con definiciones orientadas al negocio.
- Hadoop Shims: Herramienta para Big Data de Pentaho.

*Herramientas que se van a utilizar para la realización del proyecto.

Para nutrir a Pentaho se ha elegido una base de datos MySQL puesto que al disponer de la independencia que proporciona Pentaho para poder utilizar diferentes tipos de orígenes de datos se ha elegido ésta porque ya se estaba familiarizado con ella, además cuenta con una herramienta visual de diseño de base de datos MySQL Workbench (<https://www.mysql.com/products/workbench/>) lo que va a permitir gestionar la base de datos de una manera fácil y centrarnos verdaderamente en el proceso de Business Intelligence.

Todo este proceso se ha realizado en una máquina con Windows 7.

1.4 Metodología empleada

Para la realización de este proyecto se ha llevado a cabo una metodología ágil iterativa afrontando las distintas partes del proyecto de manera independiente y avanzando de una manera secuencial.

El principal método de aprendizaje ha sido documentarse todo lo posible sobre el tema a solucionar y después seguir la estrategia de prueba y error para alcanzar una solución funcional. Así como el análisis de ejemplos proporcionados en la Documentación de Pentaho Community [14] en los cuales se puede encontrar información relevante y concreta de todo lo relacionado con las distintas funcionalidades de la herramienta.

A partir de este punto, en los capítulos 2, 3, 4 y 5 se explican las partes del proceso completo de Business Intelligence, las cuales están divididas en: Instalación y configuración de Pentaho Community, Proceso ETL y creación del cubo OLAP, Construcción del Dashboard y Elaboración de Reportes. Para explicarlas se ha decidido seguir un sistema de aprendizaje tutelado mediante la realización de prácticas supervisadas en las que se establecen los detalles para poder realizar cada una de estas partes del proceso de Business Intelligence.

En las distintas partes se aprecia con capturas de pantalla datos claves de cada uno de los pasos con el objetivo de ofrecer una mayor información y que sea así mucho más sencillo para la persona que lo realiza.

2 Instalación y configuración de Pentaho Community.

2.1 Descarga y configuración de componentes

2.1.1 Java Development Kit

Cabe destacar que todos los componentes que componen Pentaho están desarrollados bajo la plataforma Java. Puesto que lo primero que debemos hacer es descargarnos la versión de java compatible con Pentaho Community más reciente, Pentaho 5.3 Community Edition no funciona con java 8, por ello, en este caso descargaremos la versión más reciente de [Java Development Kit 7](#) [15].

2.1.2 Configuración de las variables de entorno

En segundo lugar, debemos configurar las variables de entorno necesarias para que la aplicación funcione correctamente. Para ello realizaremos los siguientes pasos:

1. Pulsaremos con el botón izquierdo del ratón en el menú de inicio, después pulsaremos con el botón derecho del ratón en “equipo”, y seguido en propiedades. La Ilustración 1 muestra este proceso.

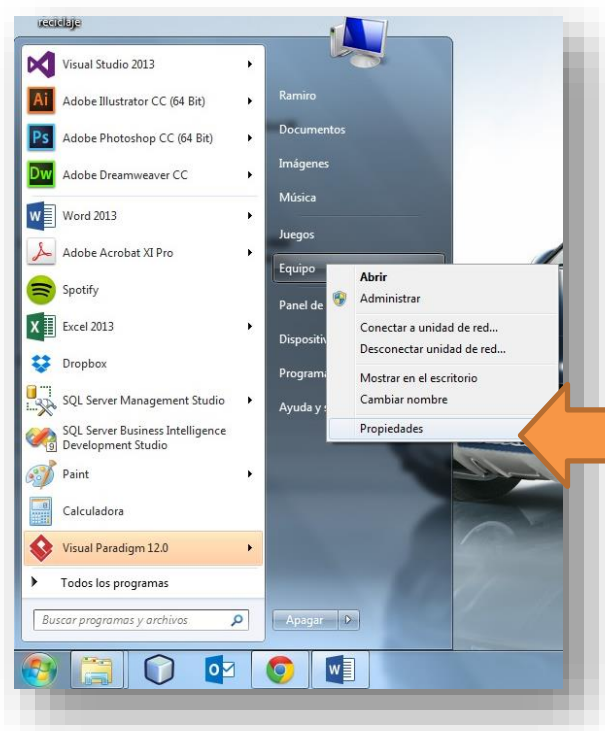


Ilustración 1: Inicio Equipo (clic derecho), Propiedades

2. Situamos el ratón sobre “Configuración avanzada del sistema” y apretamos el botón izquierdo (Ilustración 2).

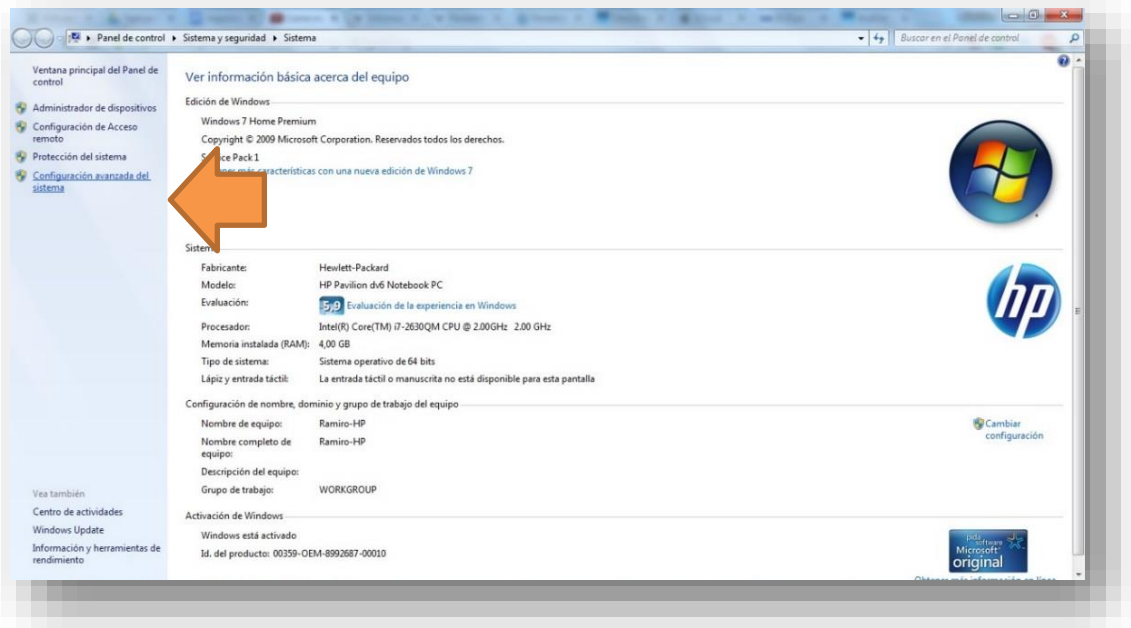


Ilustración 2: Panel de control, Configuración avanzada del sistema

3. Pulsamos con el botón izquierdo del ratón sobre “Variables de entorno de Windows”, tal y como se aprecia en la Ilustración 3.

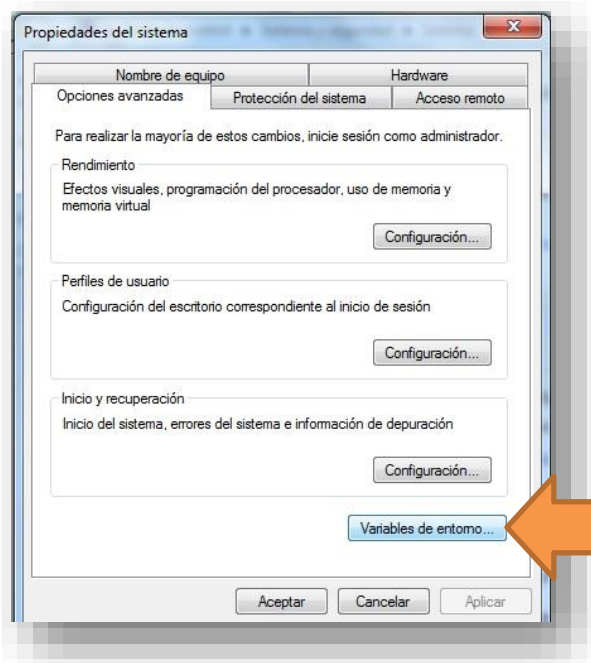


Ilustración 3: Variables de entorno

4. Editamos la variable “path” añadiendo la ruta en la que anteriormente hemos instalado el JDK.

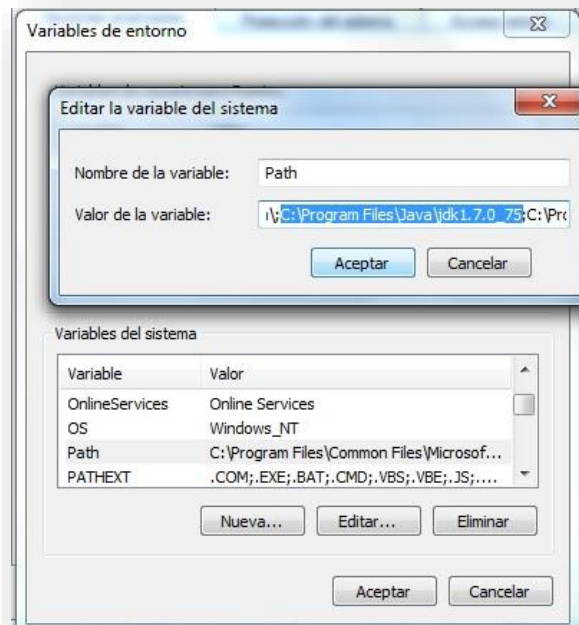


Ilustración 4: Path

5. Añadimos una nueva variable llamada “PENTAHO_JAVA_HOME”, añadiéndole como ruta, nuevamente, la del JDK.

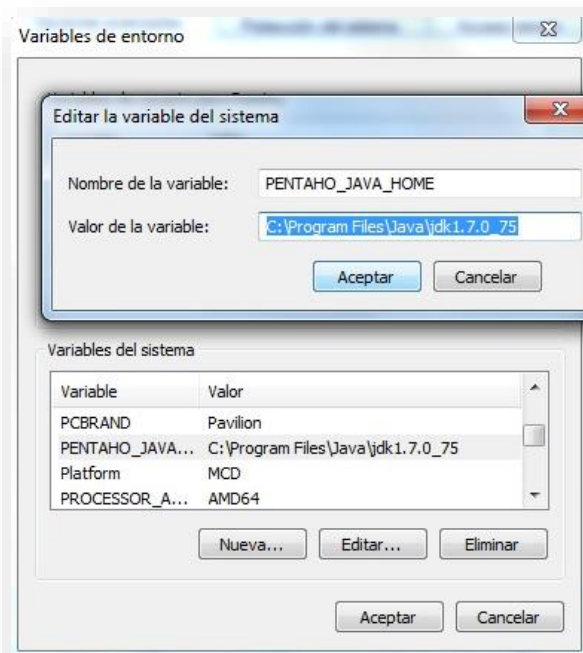


Ilustración 5: PENTAHO_JAVA_HOME

Una vez completados estos pasos ya tenemos nuestro equipo perfectamente configurado para poder utilizar el BI Server de Pentaho.

2.1.3 MySQL

Para nutrir Pentaho BI server necesitaremos un origen de datos. En nuestro caso vamos a utilizar el sistema gestor de bases de datos MySQL [16] (<https://www.mysql.com/>), y lo gestionaremos con MySQL Workbench puesto que es muy sencillo e intuitivo hacerlo con esta herramienta. Crearemos una base de datos en la que posteriormente introduciremos los datos a analizar, en concreto, y para hacerlo más fácil, a esta base de datos todos le pondremos el mismo nombre: “comercio” y las credenciales de acceso serán las siguientes usuario: “root”, password: “password”, siendo estos los valores que aparecen en MySQL. Por supuesto, su url será: localhost y el puerto, el que tiene por defecto MySQL, es decir, 3306.

2.1.4 Pentaho BI Server

Descargamos Pentaho BI server (Business Analytics Platform) de la página de Pentaho Community [2] (<http://community.pentaho.com/>), descomprimos el archivo en una carpeta creada en la raíz del disco local “D:” llamada “PENTAHO”, esto lo haremos simplemente por organización ya que en esta carpeta alojaremos todas las herramientas de Pentaho.

Una vez finalizado el proceso de descompresión, tendremos todo preparado para poder iniciar el servidor de BI de Pentaho. Para ello solo debemos ejecutar el archivo que se encuentra dentro de la carpeta “biserver-ce”, llamado “start-pentaho.bat”, con el que iniciaremos el servidor Tomcat sobre el que se aloja Pentaho y accedemos desde el navegador introduciendo la url “localhost:puerto”.

En este apartado no vamos a generar ningún nuevo rol de usuario, simplemente utilizaremos el rol de administrador que trae por defecto Pentaho, al cual podremos acceder con las siguientes credenciales, usuario: “admin”, contraseña: “password”. Pero cabe destacar que Pentaho permite gestionar los roles de usuario, incluyendo los permisos y vistas a los que tienen acceso para poder configurar quién queremos que vea qué.

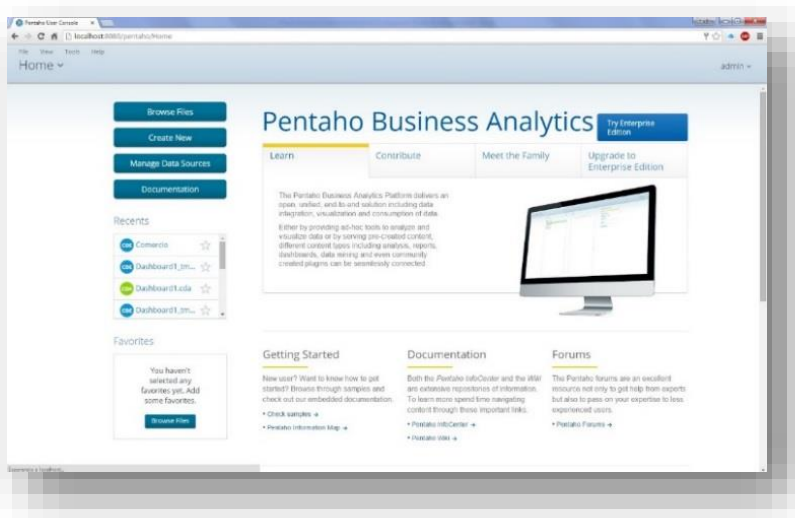


Ilustración 6: Pantalla inicial Pentaho

2.1.5 Data Sources

Una vez llegado a este punto, tendremos instalados Pentaho BI server y la base de datos MySQL, ahora bien, es necesario añadir el origen de datos en Pentaho, para que éste pueda acceder a los datos de nuestra base de datos. Para ello seguiremos los siguientes pasos:

1. Accedemos a *File, Manage Data Sources*, tal y como se aprecia en la ilustración 7.



Ilustración 7: Manage Data Sources

2. Nos aparecerá una ventana en la que veremos todos los orígenes de datos incluidos en Pentaho hasta el momento. En nuestro caso aparecerán los de los ejemplos. Como se aprecia en la Ilustración 8, accedemos a *New Data Source*.

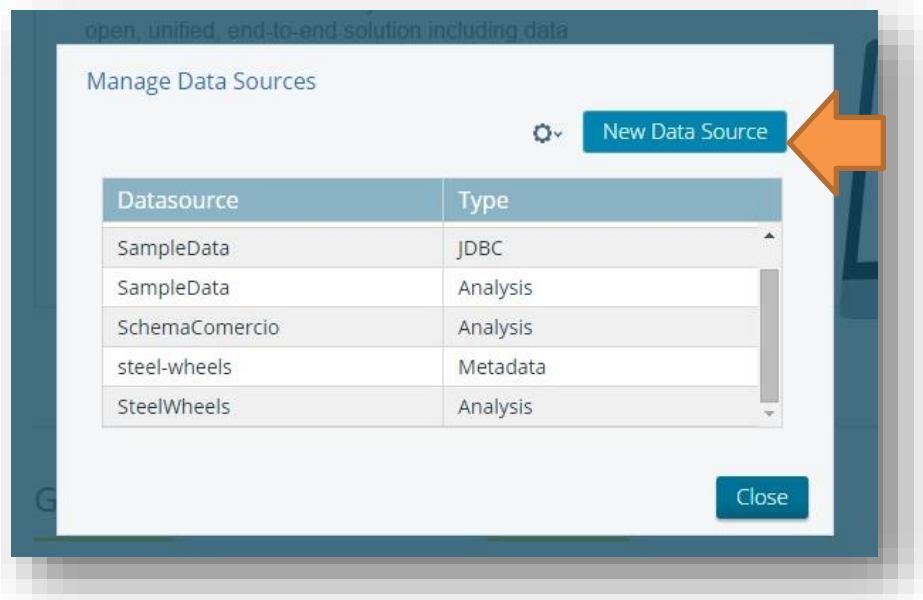


Ilustración 8: New Data Source 1

3. Como se muestra en la Ilustración 9, seleccionamos como *Source Type: Database Table(s)*, posteriormente, y como se detallará en el siguiente paso, añadimos una nueva conexión pulsando “+”.

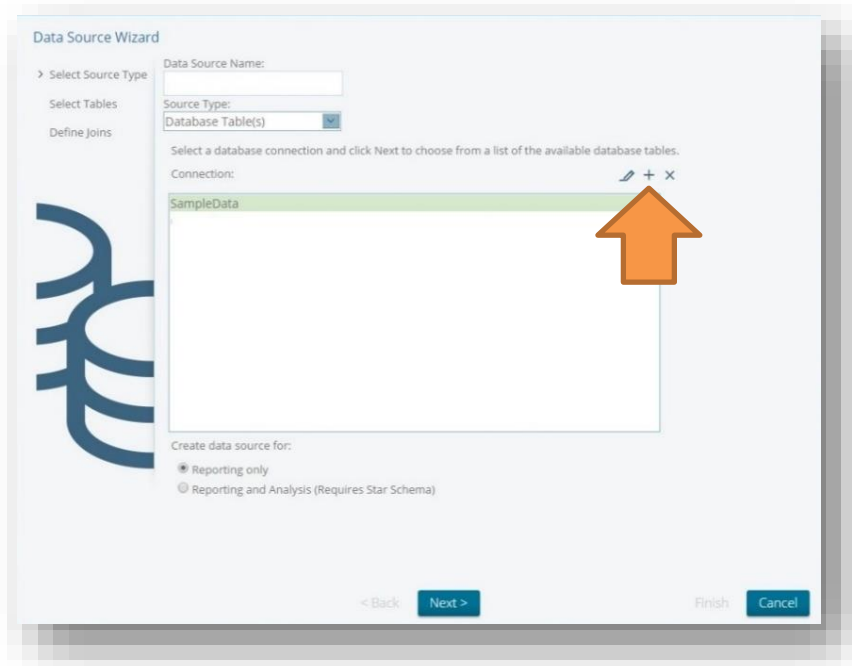


Ilustración 9: New Data Source 2

4. Como se aprecia en la Ilustración 10, seleccionamos nuestro tipo de base de datos, MySQL, e introducimos todo lo requerido para establecer la conexión.

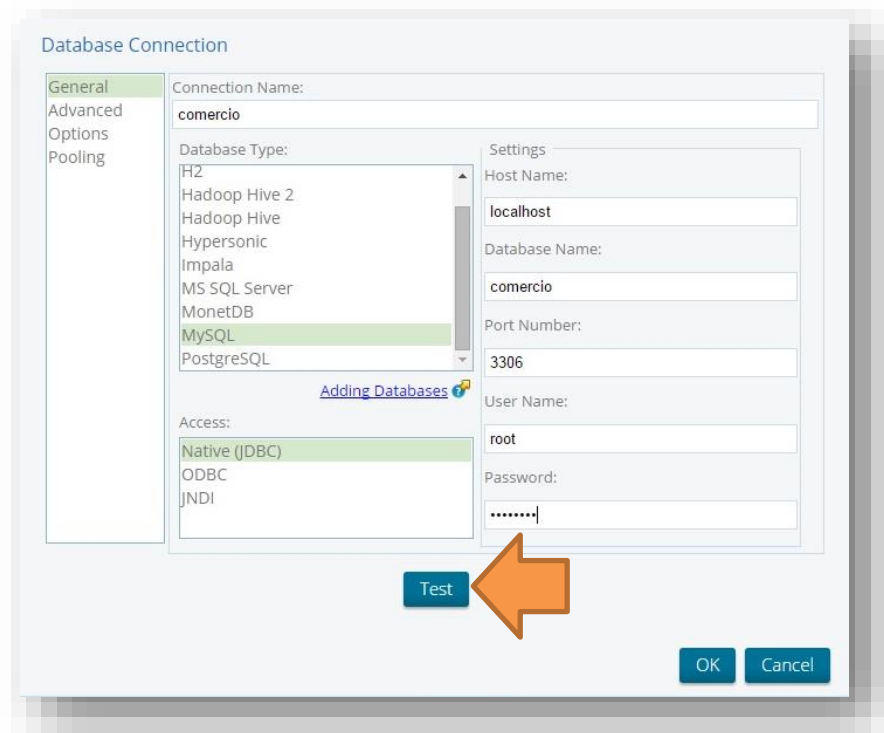


Ilustración 10: Credenciales MySQL

5. Una vez hecho el Test y comprobado que es satisfactorio (Ilustración 11), haremos clic en OK y ya tendremos establecida la conexión entre Pentaho y la base de datos comercio que tenemos en MySQL.

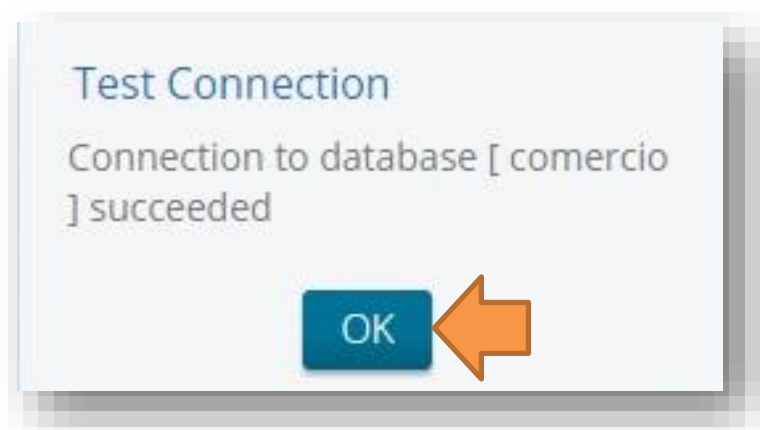


Ilustración 11: Test Data Source

6. Como se observa en la Ilustración 12, ya disponemos de la conexión “comercio” por lo que ya podemos salir del asistente.

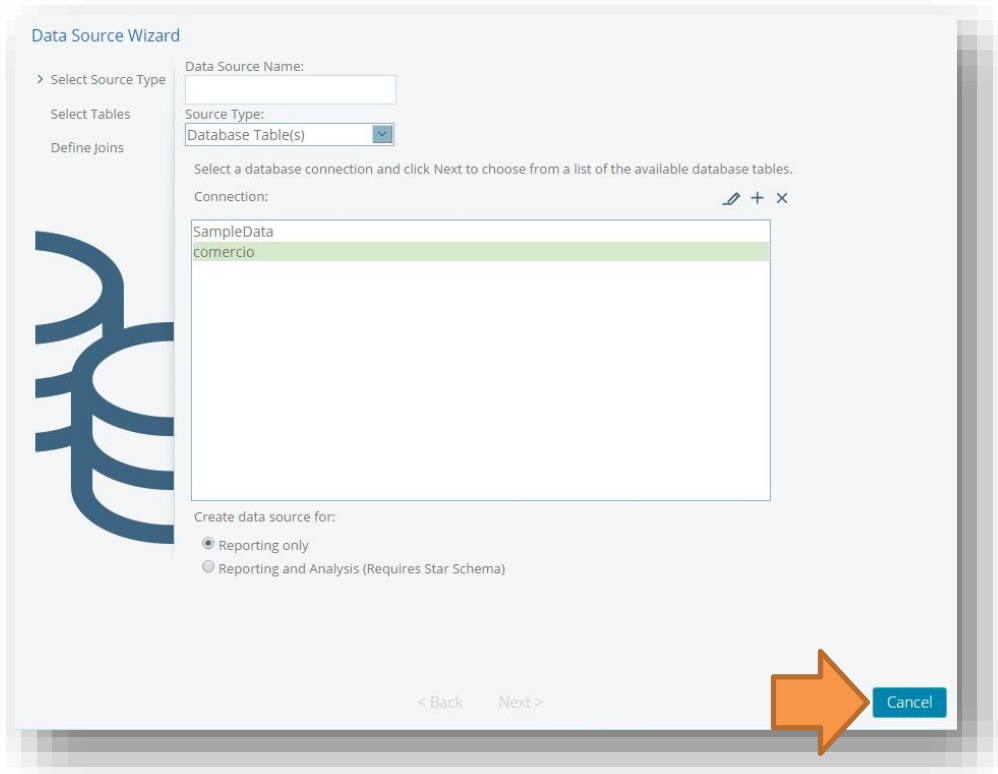


Ilustración 12 Data Source OK 1

7. Apreciamos en la Ilustración 13 que ya tenemos creado nuestro *Data Source* “comercio”.

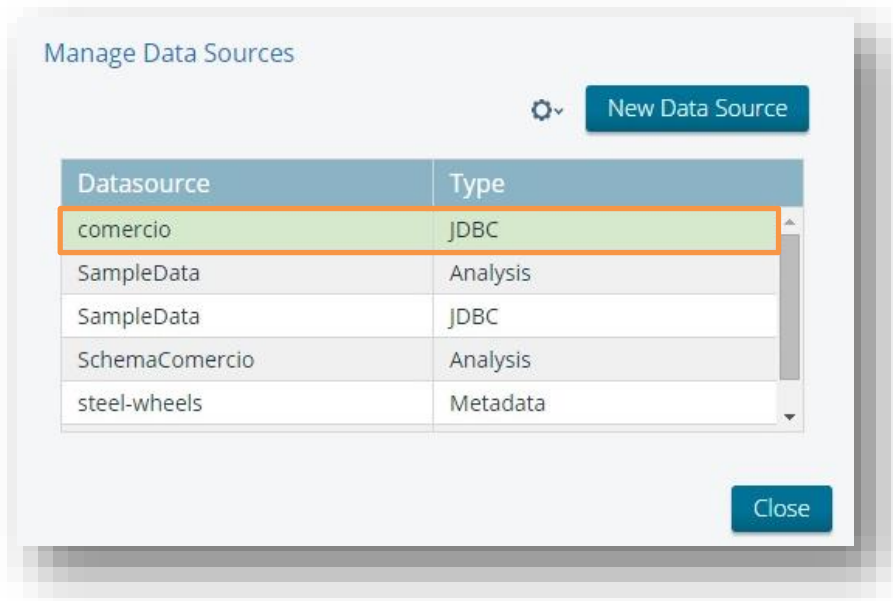


Ilustración 13: Data Source OK 2

2.2 Resto de componentes de Pentaho

Para el resto de componentes no es necesario configurar nada a nivel de variables locales. Simplemente descargamos el archivo comprimido desde la página web de Pentaho Community [2] y lo descomprimos en la carpeta “PENTAHO” que tenemos creada en la raíz de nuestro disco local “D:”. Dentro de esa carpeta descomprimida encontraremos un archivo con el nombre de la aplicación en cuestión y ejecutándolo directamente se iniciará dicha aplicación.

Para este proyecto es necesario contar, además de con Pentaho BI Server que ya tenemos satisfactoriamente instalado y configurado, con:

- Data Integration
- Schema Workbench
- Report Designer

2.3 Instalación del Marketplace de Pentaho

Para poder utilizar el Marketplace de Pentaho, y así poder instalar plugins con nuevas funcionalidades, debemos descargarlo de la página web de Pentaho Community [2] (<http://community.pentaho.com/>) y descomprimir el archivo “.zip” descargado en la ruta “C:\Pentaho\biserver-ce\pentaho-solutions\system”. Una vez hecho esto, reiniciamos el servidor y ya tendremos disponible el Marketplace. La Ilustración 14 muestra la pestaña de acceso al Marketplace.

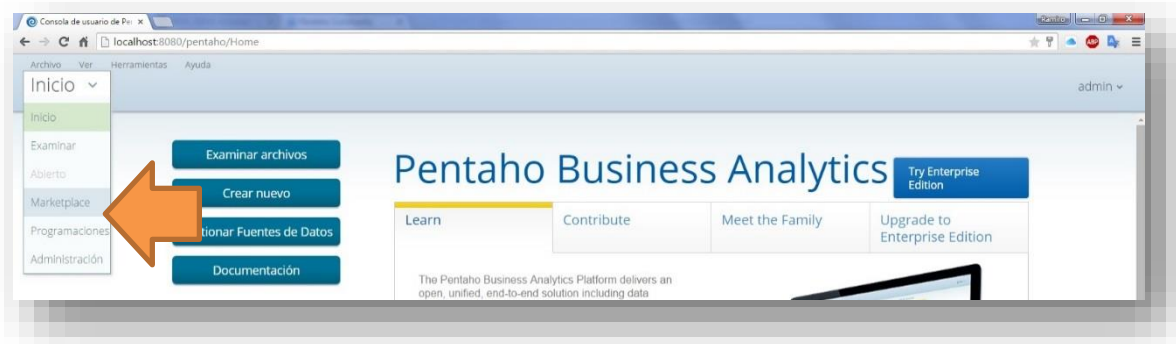


Ilustración 14: Marketplace

2.4 Instalación de Saiku Analytics 2.6 (DEPRECIATED)

Vamos a instalar la versión 2.6 de Saiku Analytics ya que nos va a permitir hacer consultas de una manera mucho más intuitiva y gráfica. Instalaremos la versión 2.6 puesto que, aunque no es la más reciente, sí que es la que no incorpora ningún error para la utilidad que le queremos dar.

Buscamos el plugin Saiku Analytics como se ve en la ilustración 15.

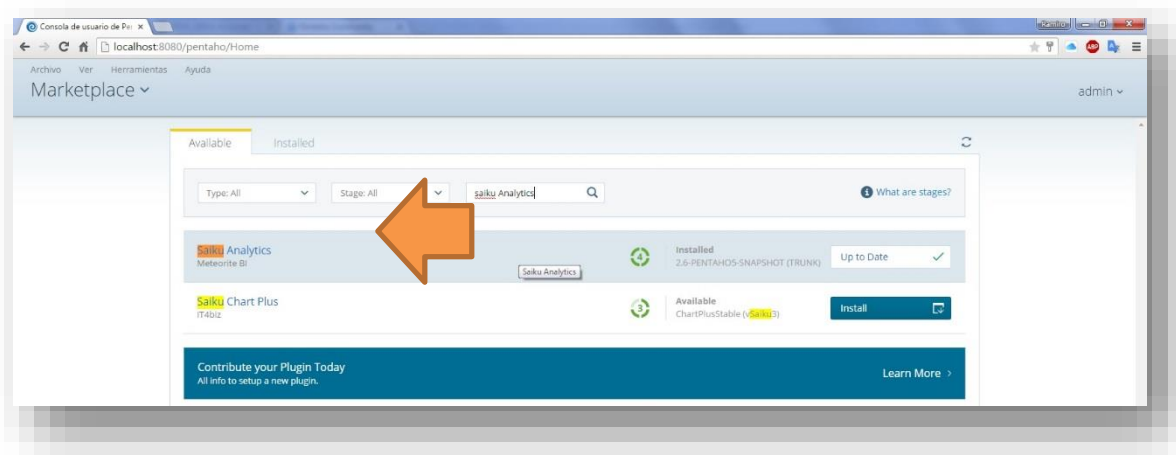


Ilustración 15: Saiku Analytics

Posteriormente seleccionamos la versión 2.6 de entre las disponibles y la instalamos, tal y como se muestra en la Ilustración 16. Después habrá que reiniciar Pentaho y ya podremos utilizar el plugin sin problema alguno.



Ilustración 16: Saiku Analytics Versión 2.6

2.5 Conclusión

Una vez realizado todo esto ya disponemos del enlace necesario para que Pentaho pueda acceder a los datos de nuestra base de datos “comercio”. En el siguiente apartado procederemos a crear las tablas de nuestra base de datos y poblarla con los datos del Comercio Exterior recogidos de la página web del INE [1] y comprendidos entre 1994 y 2012.

3 Proceso ETL y creación del cubo OLAP.

3.1 Proceso ETL

Una vez iniciada la aplicación, ejecutando el archivo “Spoon.bat” que se encuentra en la carpeta “D:\data-integration” nos encontraremos, tal y como se aprecia en la ilustración 17, con una pantalla inicial en la que aparece una presentación y una serie de documentación sobre la aplicación. Esta aplicación también se denomina Spoon.

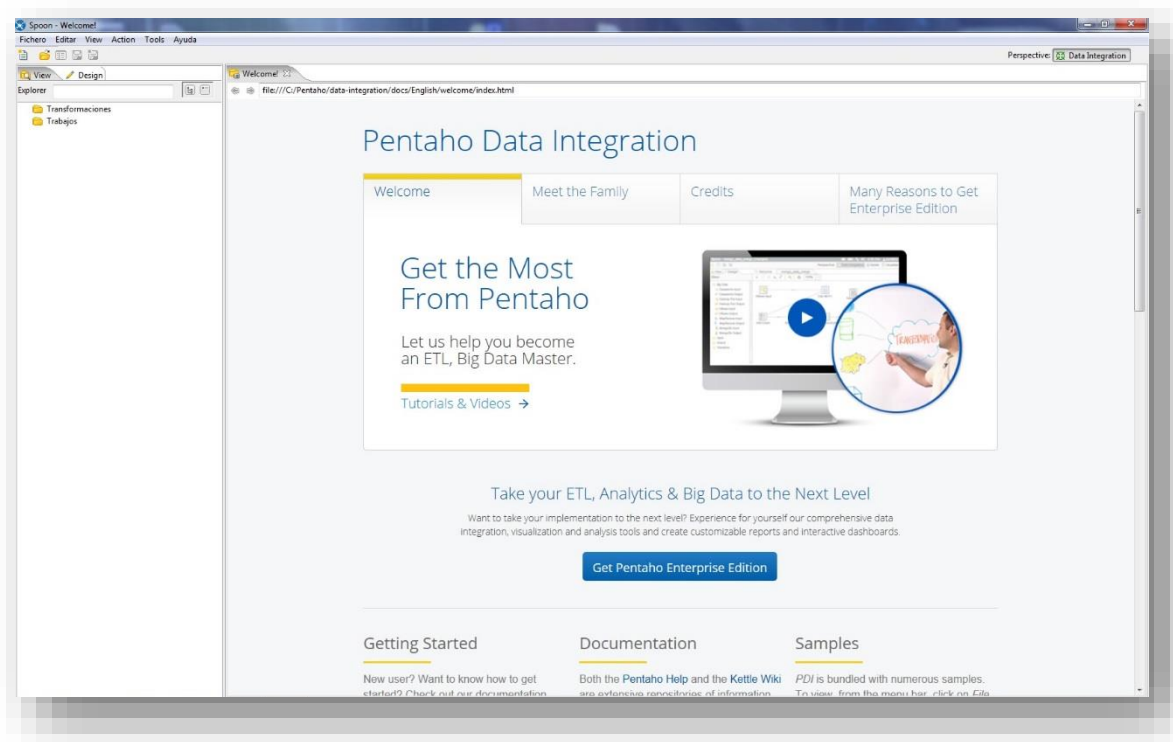


Ilustración 17: Pantalla inicial Pentaho Data Integration

Para la realización de este apartado contaremos con los datos disponibles en los documentos Excel “ImportacionesYExportacionesCLAVEVALOR.xlsx”, “Año.xlsx”, “Sector.xlsx” y “Tipo.xlsx” en los que se encuentran los datos de las importaciones y exportaciones españolas desde el año 1994 hasta el año 2012 extraídos del INE [1]. Con ellos vamos a construir un cubo OLAP con una topología en estrella que la formarán una tabla de hechos y las tablas de las dimensiones, relacionadas con la tabla de hechos a través de una clave foránea.

3.1.1 Nueva transformación

Para realizar nuestro proceso ETL, tal y como podemos ver en la ilustración 18, crearemos una nueva transformación en la que vamos a explicar todo lo necesario para introducir los datos en nuestra base de datos MySQL de forma correcta.

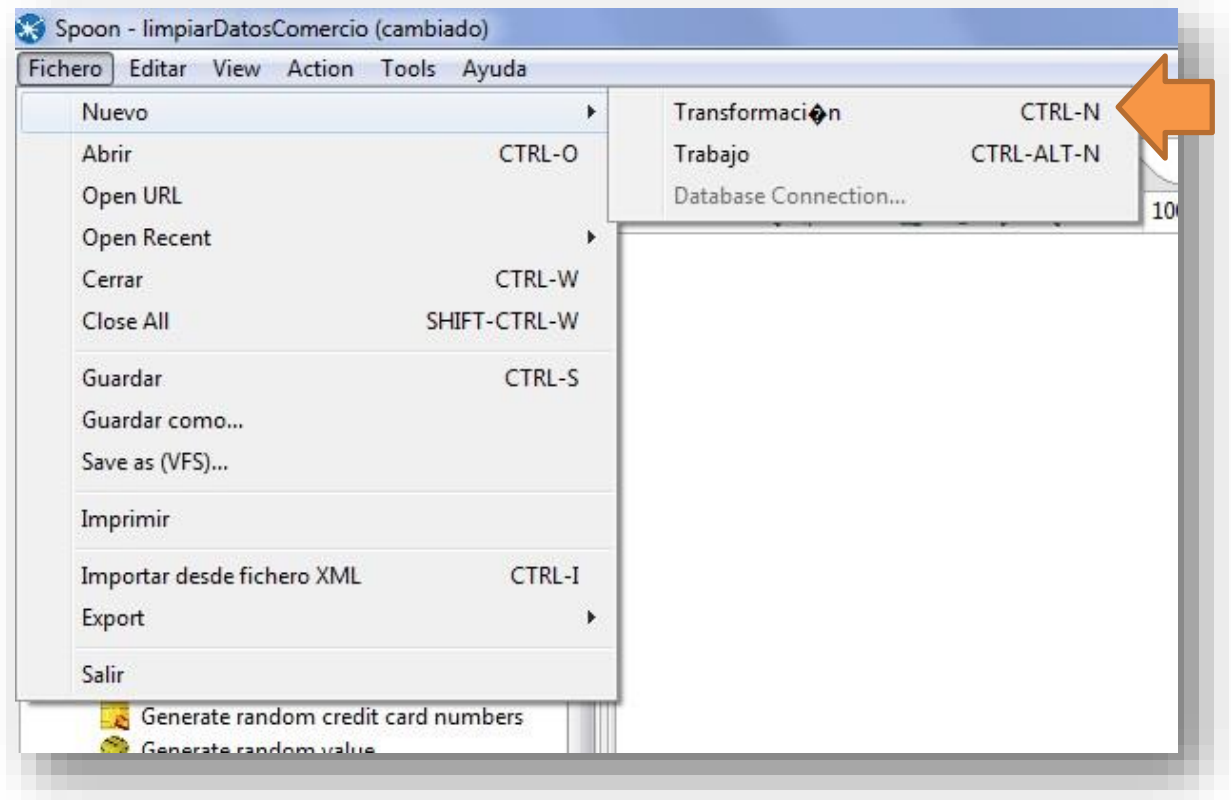


Ilustración 18: Fichero, Nuevo, Transformación

3.1.2 Entrada Excel

Disponemos de unos archivos Excel en los que se encuentran los datos del comercio exterior de España (1994-2012). Por ello, seleccionaremos como entrada de datos en la nueva transformación “Entrada Excel”, tal y como se muestra en la ilustración 19.

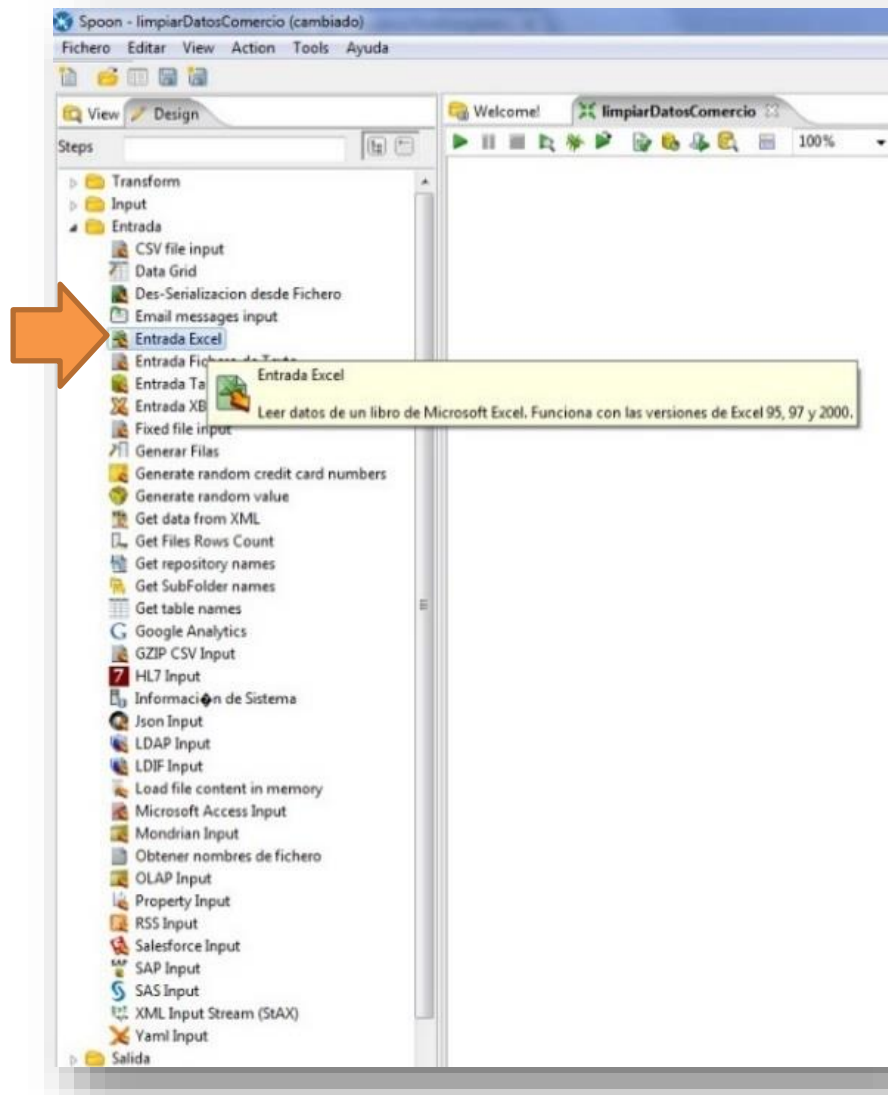


Ilustración 19: Entrada Excel 1

Una vez agregado el icono de “Entrada Excel”, haciendo doble clic sobre él aparecen las propiedades de dicho icono en donde configuraremos las opciones que se indican a continuación.

Tal y como vemos en la ilustración 20, añadimos el archivo del que queremos obtener los datos.

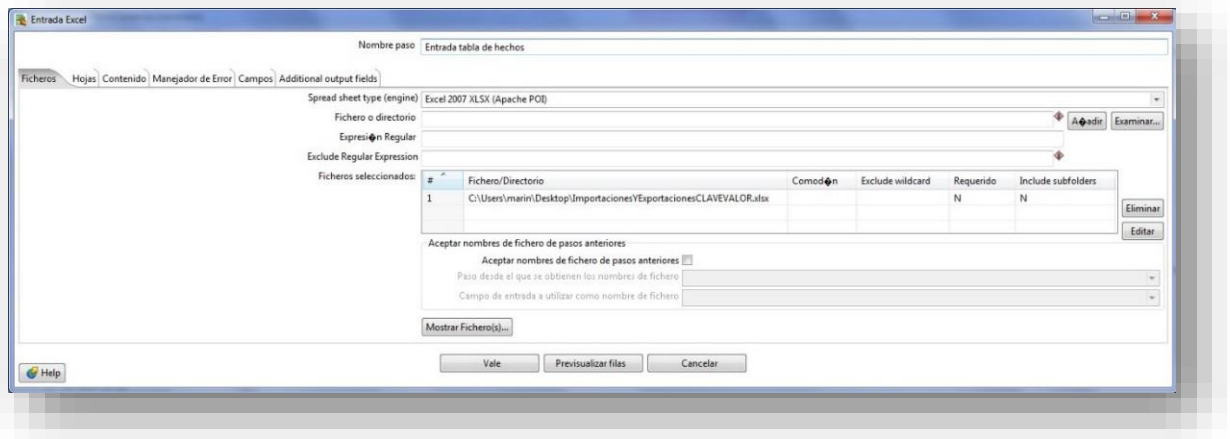


Ilustración 20: Entrada Excel 2

Posteriormente seleccionamos las hojas como se indica en la ilustración 21.

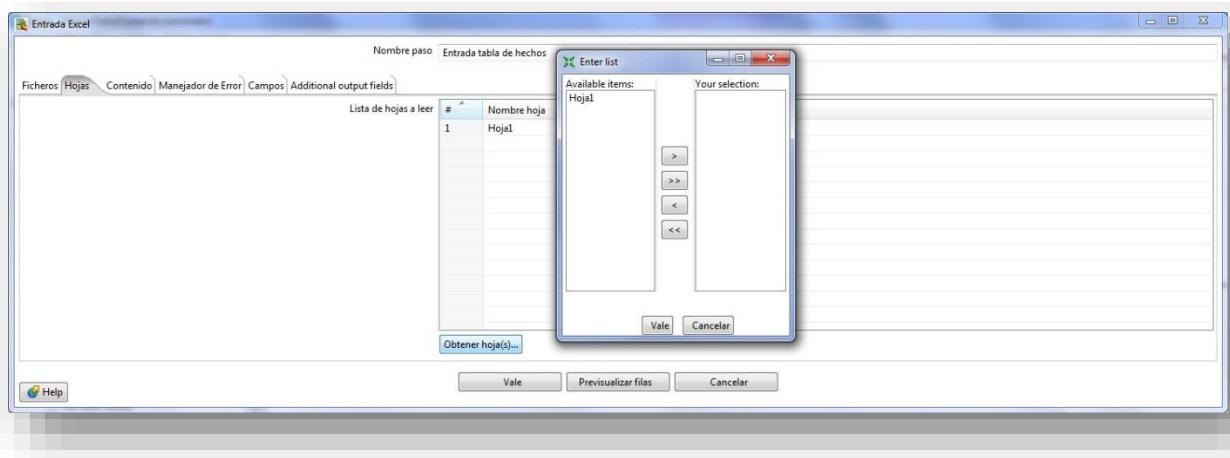


Ilustración 21: Entrada Excel 3

Por último, como vemos en la ilustración 22, obtenemos los campos del archivo Excel y seleccionamos el tipo de dato que mejor se adapte a nuestras necesidades.

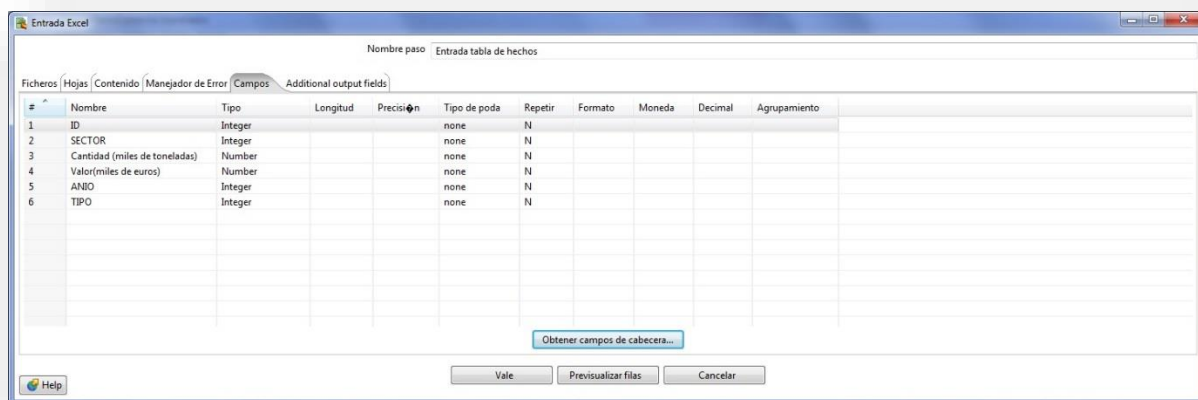


Ilustración 22: Entrada Excel 4

3.1.3 Salida tabla

En nuestro caso concreto, como ya disponemos de los datos en un formato adecuado, no vamos a realizar ninguna transformación. Así que directamente agregamos los datos a una tabla, seleccionando “Salida tabla”. Nos aparecerá un menú como el de la Ilustración 23 y agregaremos una nueva conexión.

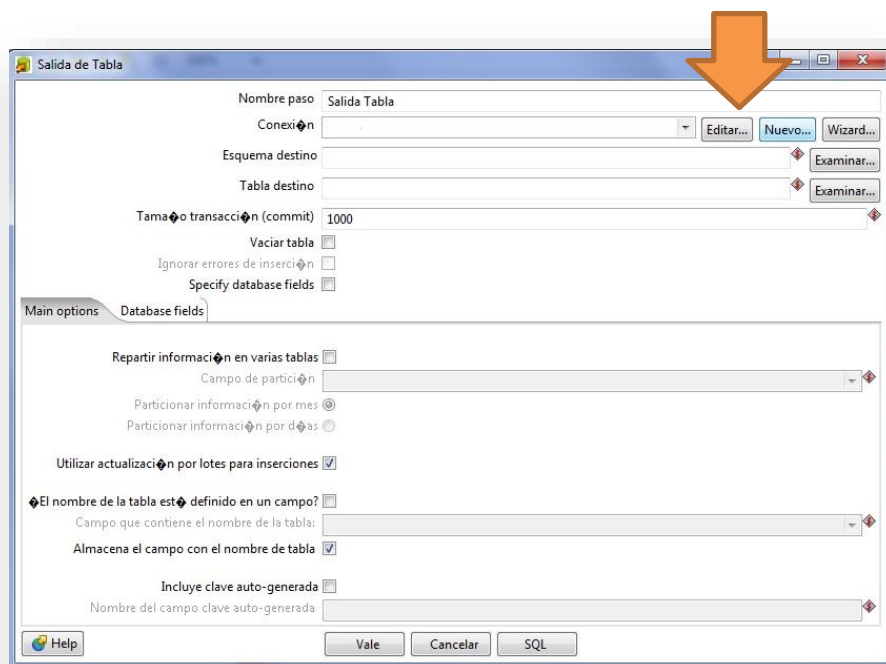


Ilustración 23: Salida Tabla

A continuación introduciremos los datos de conexión de nuestra base de datos MySQL. Para ello, primeramente, debemos añadir la librería de Java a la carpeta “\data-integration\lib”, que la podemos descargar de la página web de conectores de MySQL [17]. (<http://dev.mysql.com/downloads/connector/j/>) (Ver ilustración 24).

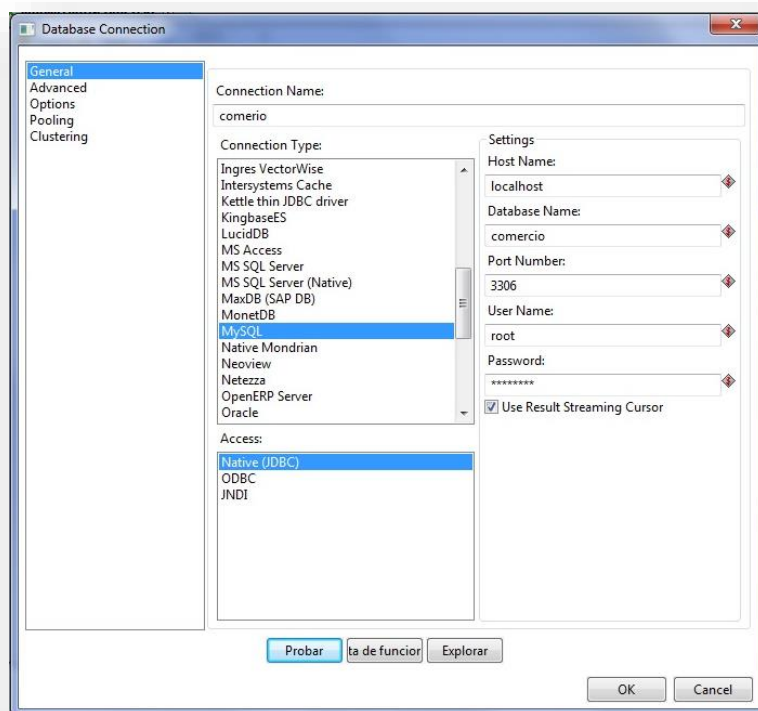


Ilustración 24: Database Connection ETL

Después de comprobar que la conexión se ha establecido correctamente, podemos continuar (ver Ilustración 25).

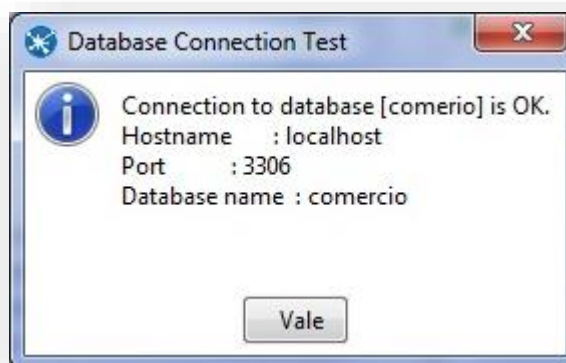


Ilustración 25: Database Connection Test ETL

Una vez tenemos configurada la conexión escribimos un nombre de tabla y procedemos a pulsar sobre el botón “SQL”, tal y como se ve en la Ilustración 26, que automáticamente generará un script con el nombre de la tabla a crear y los campos con su tipo de dato correspondiente. Si ejecutamos dicho script se crearán las tablas correspondientes en nuestra base de datos.

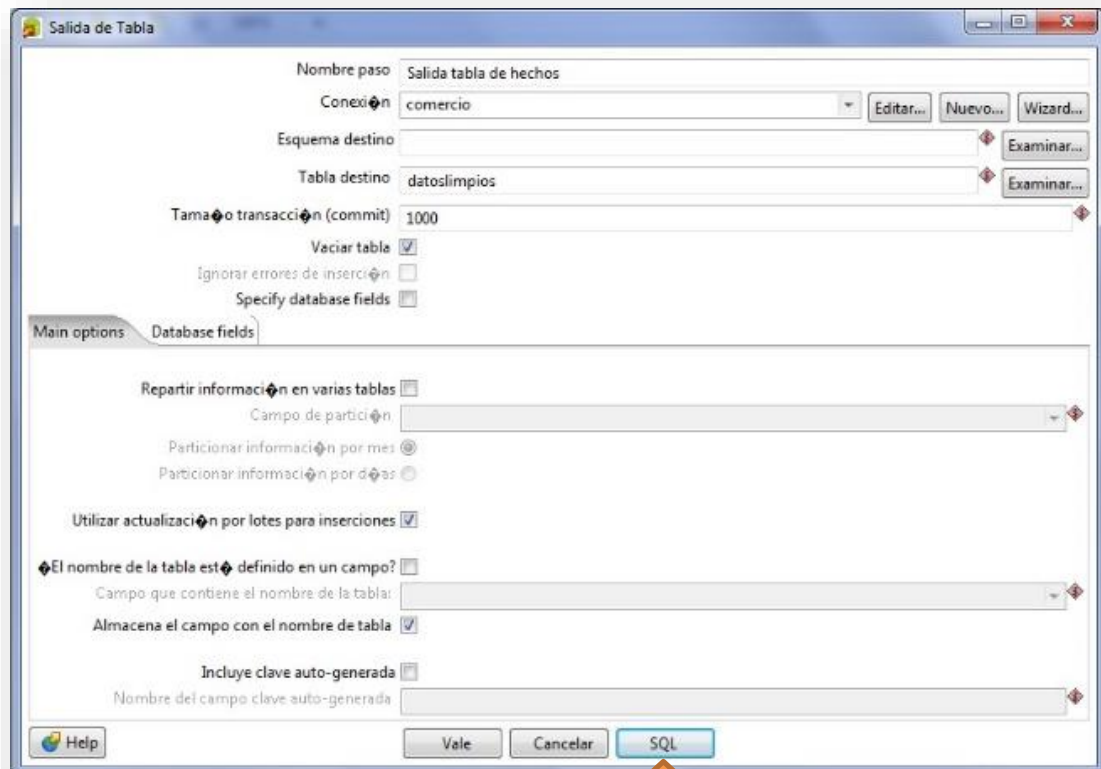


Ilustración 26: Salida Tabla 2

3.1.4 Vista general del ETL

Una vez realizado el proceso para una de las tablas, procederemos de la misma manera para hacerlo con el resto de ellas, obteniendo así el esquema que se muestra en la Ilustración 27.

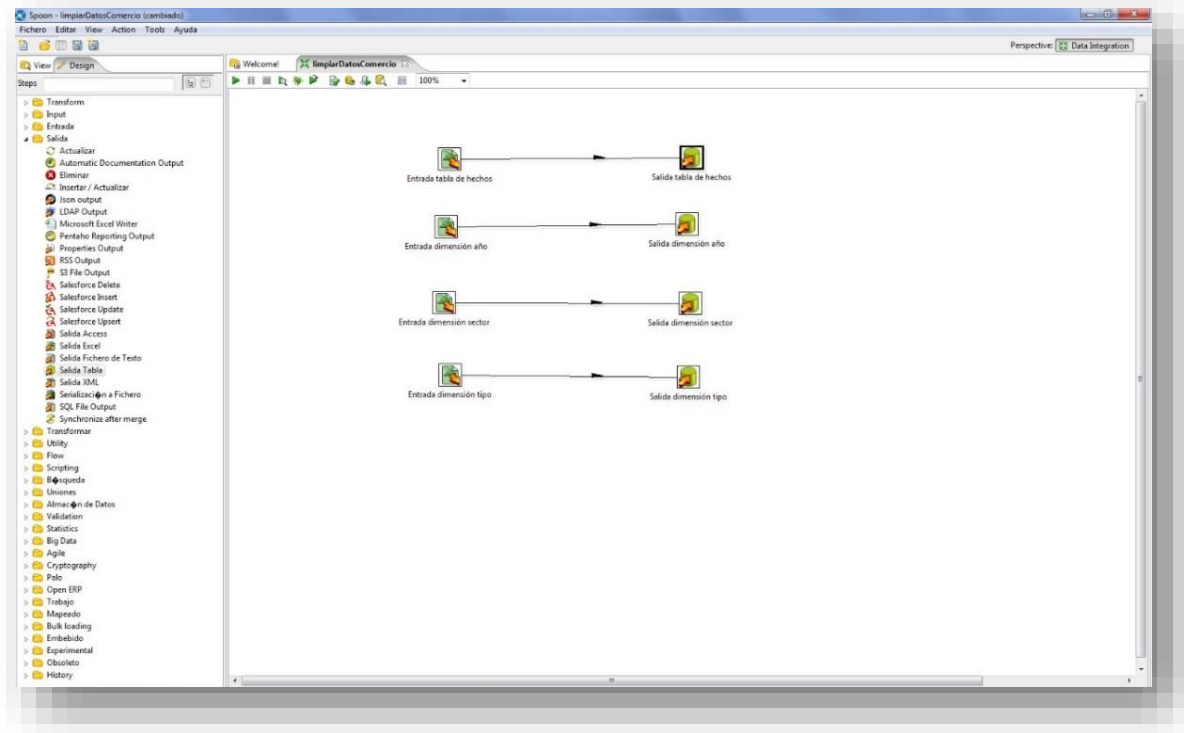


Ilustración 27: Vista General ETL

Después de ello podremos darle al botón play y posteriormente a ejecutar para llevar a cabo la carga de datos, tal y como se aprecia en la ilustración 28. Después de esto, todos los datos serán añadidos a las tablas de nuestra base de datos MySQL.

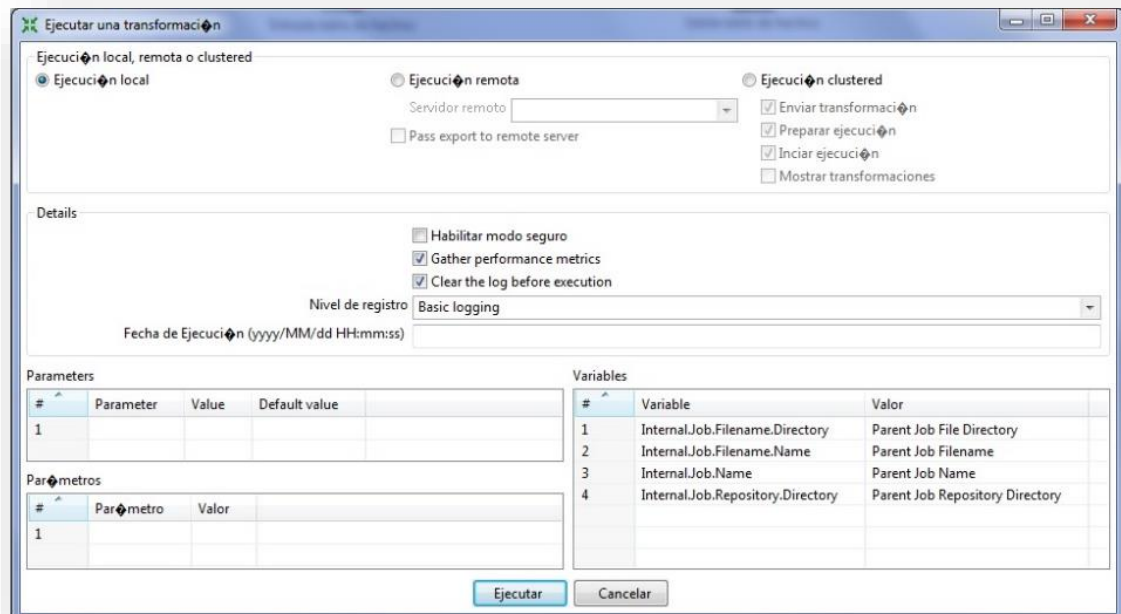


Ilustración 28: Ejecutar transformación

Por último, desde MySQL Workbench asignaremos las claves principales y foráneas a todas las tablas. Cabe destacar que éste es un paso muy importante puesto que de no ser así, no podremos construir el cubo OLAP de manera correcta. Una vez hecho esto tendremos la topología en estrella que se ve en la Ilustración 29.

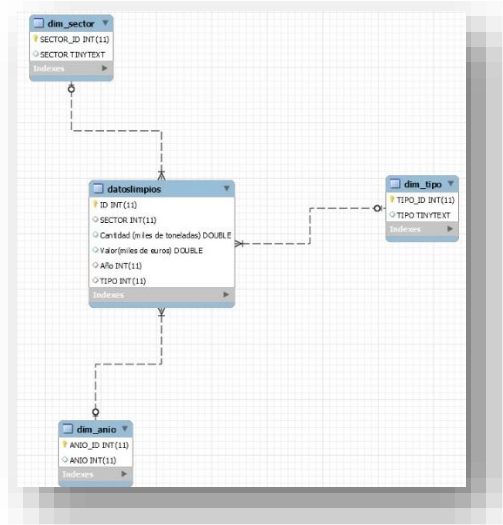


Ilustración 29: Topología en estrella

3.2 Creación de cubo OLAP con Schema Workbench

La herramienta para crear cubos de Pentaho es Schema Workbench. Lo primero que debemos hacer es establecer la conexión con nuestra base de datos MySQL, accediendo a “options” y posteriormente a “database connection” nos aparecerá un cuadro de opciones de conexión como el que se ve en la Ilustración 30.

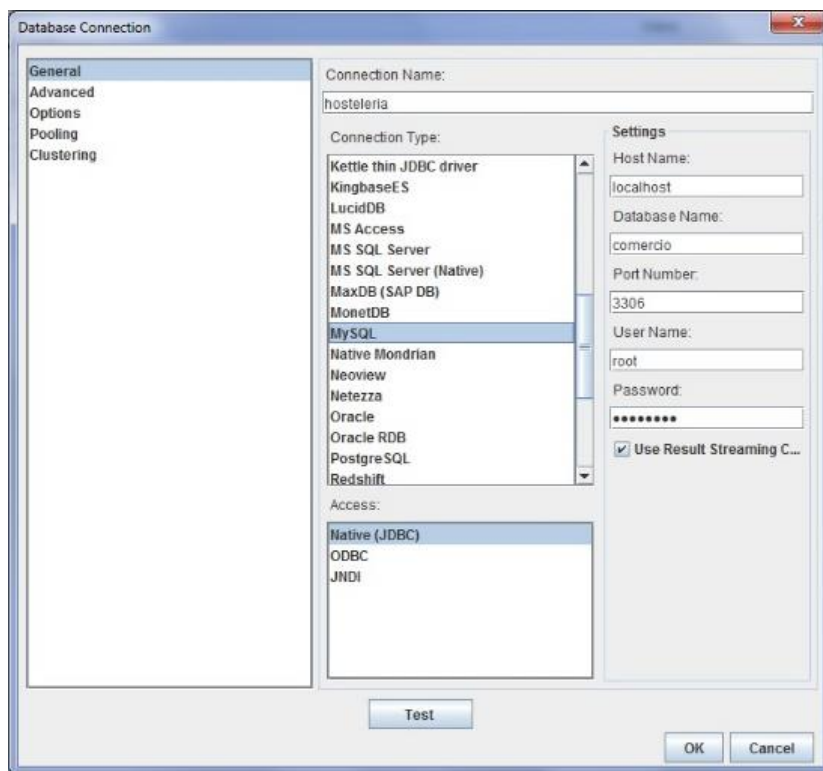


Ilustración 30: Database Connection Schema Workbench

Posteriormente crearemos un nuevo esquema, como se aprecia en la Ilustración 31.

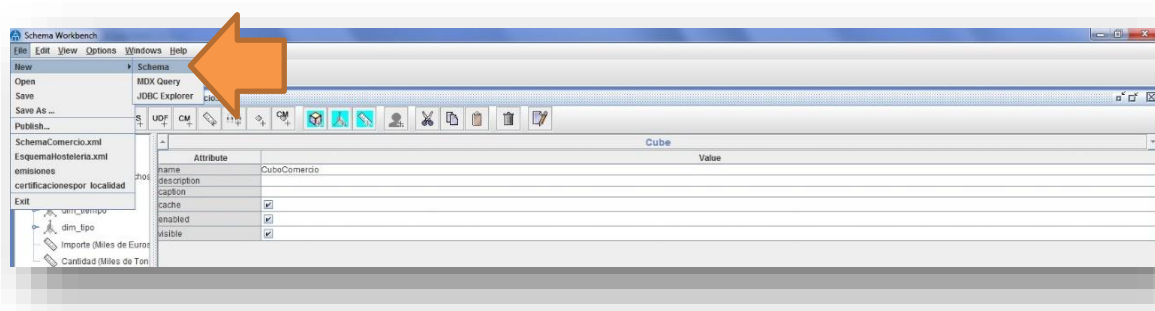


Ilustración 31: File, New, Schema

Después ya podemos añadir un nuevo cubo, tal y como se aprecia en la Ilustración 32.

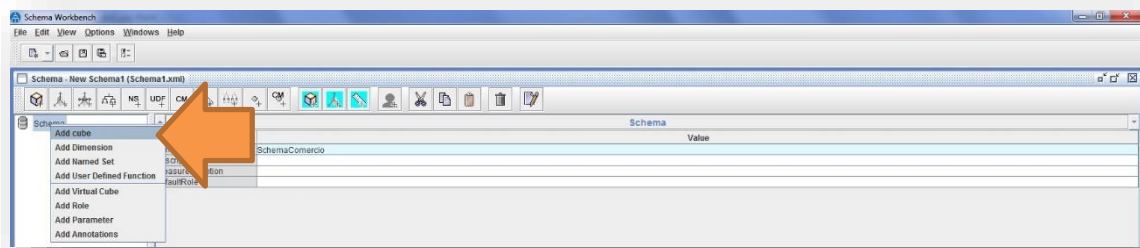


Ilustración 32: Add Cube

A continuación vamos a ir añadiendo las dimensiones al cubo. Añadimos una nueva dimensión. En este caso vamos a centrarnos en la dimensión tiempo. Para seguir una nomenclatura común, llamaremos a la dimensión “dim_nombre”. Como se ve en la Ilustración 33, añadimos una nueva dimensión.

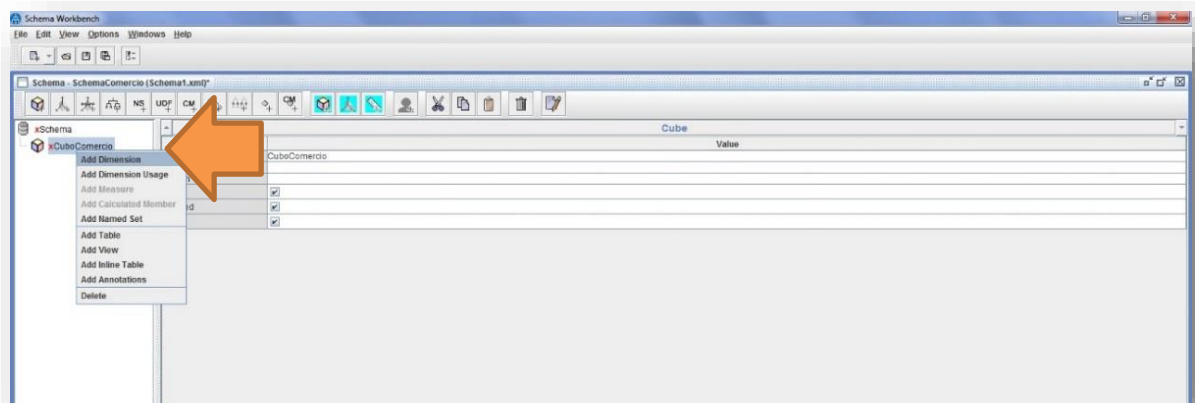


Ilustración 33: Add Dimensions

Tal y como se aprecia en la Ilustración 34, introducimos el nombre de la dimensión siguiendo la nomenclatura especificada. Especificamos cuál es la clave extranjera y el tipo de dimensión que estamos creando, en este caso, dimensión temporal.

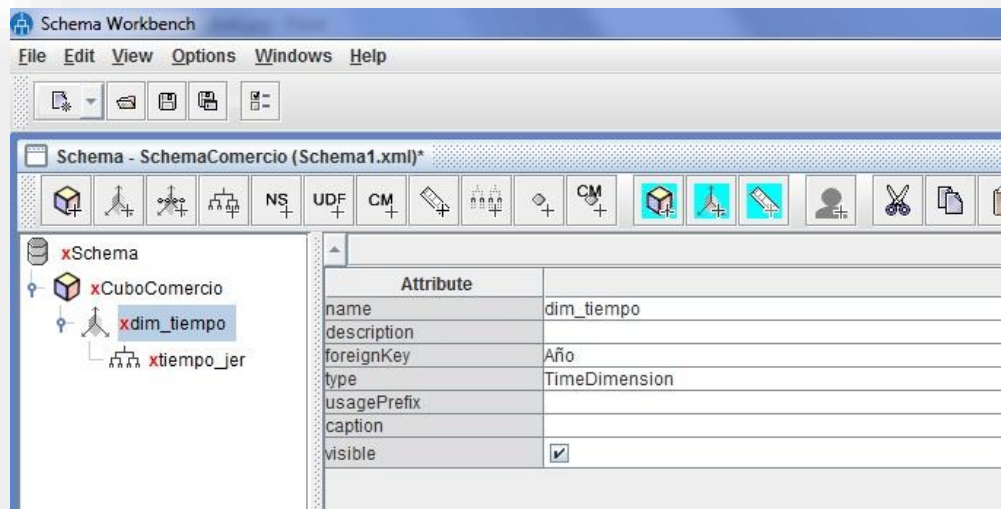


Ilustración 34: Dimensions

Desplegamos la dimensión tiempo y veremos que está creada una jerarquía. Le asignamos un nombre congruente y procedemos, como se aprecia en la Ilustración 35, a añadir una nueva tabla bajo esa jerarquía. En nuestro caso, se trata de la tabla “año”, tal y como se muestra en la Ilustración 36.

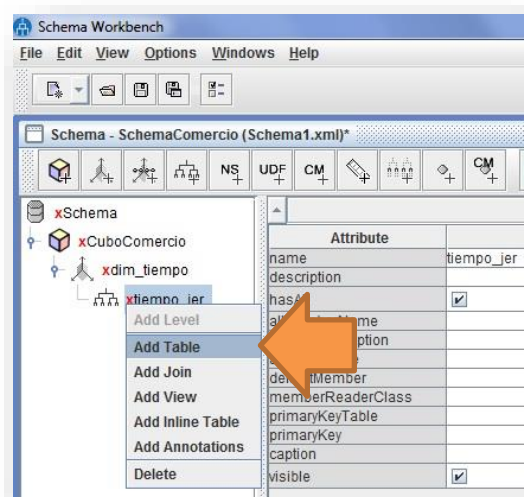


Ilustración 35: Add Table

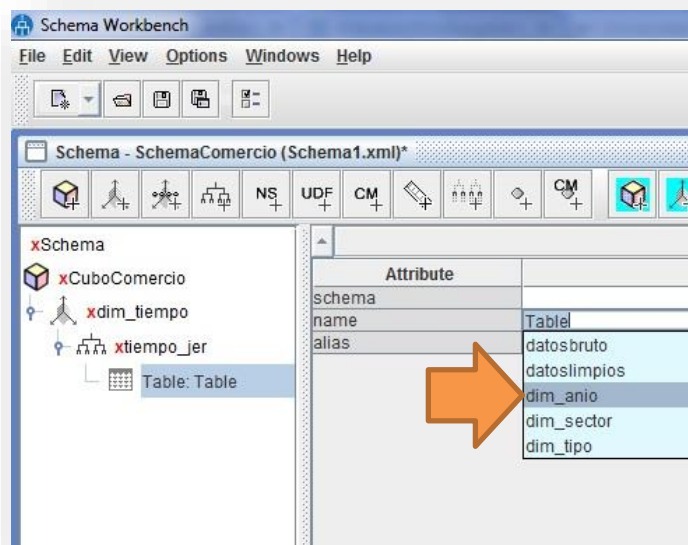


Ilustración 36: Table

Añadimos un nuevo nivel a la jerarquía (ver Ilustración 37).

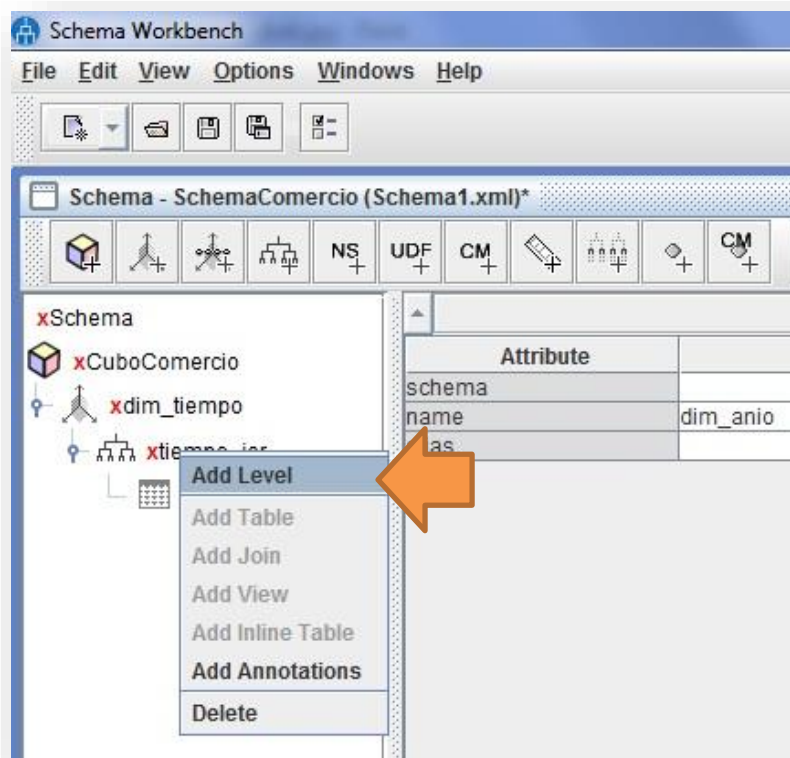


Ilustración 37: Add Level

En nuestro caso, y como se aprecia en la Ilustración 38, será el nivel año. También especificamos la columna y el nombre de la columna, y en el tipo de nivel seleccionamos TimeYears.

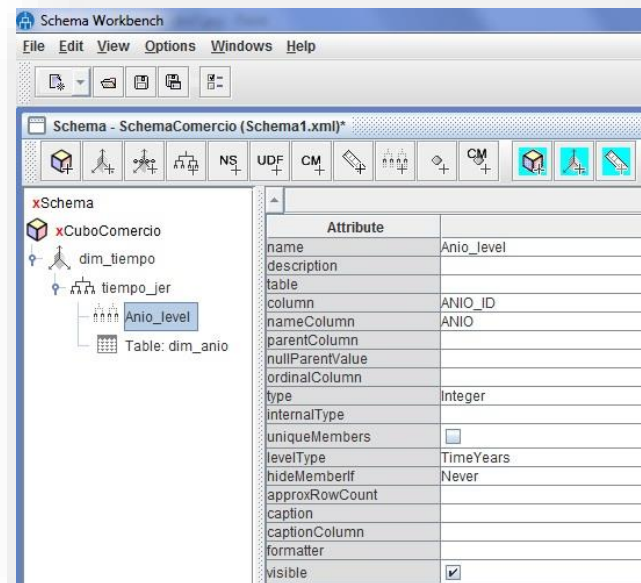


Ilustración 38: Level

Repetimos todos estos pasos con el resto de dimensiones. Añadimos una o varias medidas (*measures*) y ya tendremos creado el cubo OLAP, tal y como se muestra en la Ilustración 39.

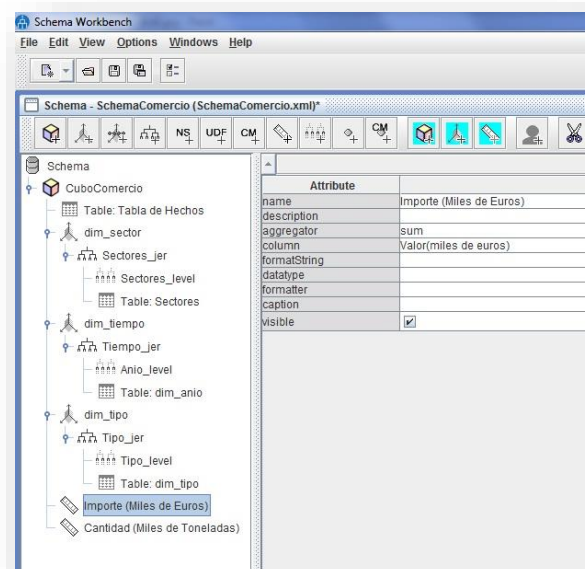


Ilustración 39: Measures

Una vez terminado el cubo lo único que debemos hacer es publicarlo introduciendo como credenciales las de administrador de Pentaho BI Server (ver Ilustración 40).

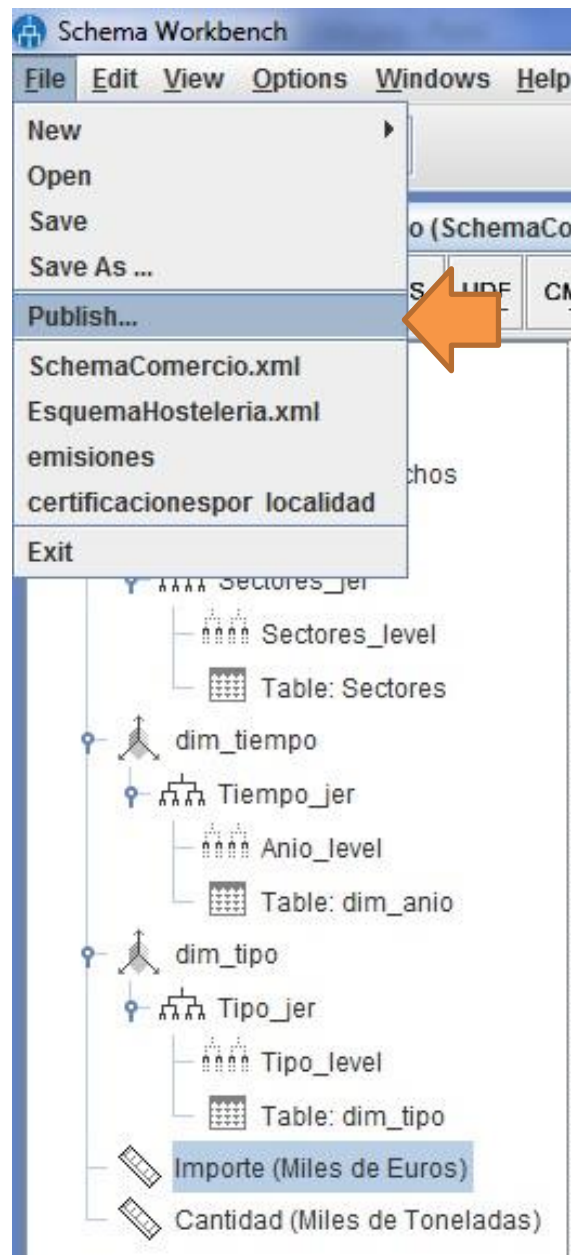


Ilustración 40: File, Publish

3.3 Visualización del cubo OLAP con JPivot View

Para asegurarnos de que hemos construido y publicado el cubo OLAP correctamente, lo podemos visualizar de una forma muy rápida y sencilla con JPivot dentro de Pentaho. Para ello vamos a la pestaña crear nuevo y allí seleccionamos JPivot View (ver Ilustración 41). Posteriormente seleccionamos el cubo sobre el que queremos generar la vista (Ilustración 42) y después nos aparecerán los datos agrupados por las dimensiones definidas (Ilustración 43).



Ilustración 41: Crear nuevo, JPivot View



Ilustración 42: New JPivot View

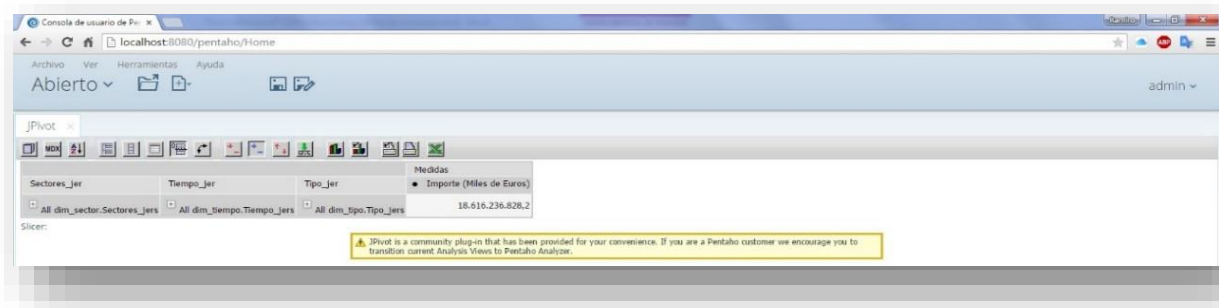


Ilustración 43: JPivot View

4 Construcción del Dashboard.

4.1 Consideraciones para construir un gráfico adecuado

Para poder mostrar y manipular los datos de una manera rápida e intuitiva vamos a construir un Dashboard con la herramienta de creación de Dashboards de Pentaho. Cabe destacar que debemos intentar que nuestro tablero de mandos sea intuitivo y fácil de manejar puesto que vamos a requerir de él en el momento de tomar una decisión y nos tiene que permitir analizar diferentes casuísticas de una manera ágil para poder hacerlo con la suficiente eficacia y poder tomar una decisión estratégica adecuada.

Por otro lado, como es lógico, deberá ser visual e incorporar gráficos apropiados. Un ejemplo de gráfico sería mostrar una evolución anual en un gráfico lineal con una escala adecuada que mantenga una proporcionalidad correcta. No sería conveniente que se utilizara un fondo de escala que ofreciera una sensación visual engañosa. Por ejemplo, vemos cómo cambia la sensación visual de la gráfica en la Ilustración 44 con los mismos datos, únicamente cambiando el fondo de escala.

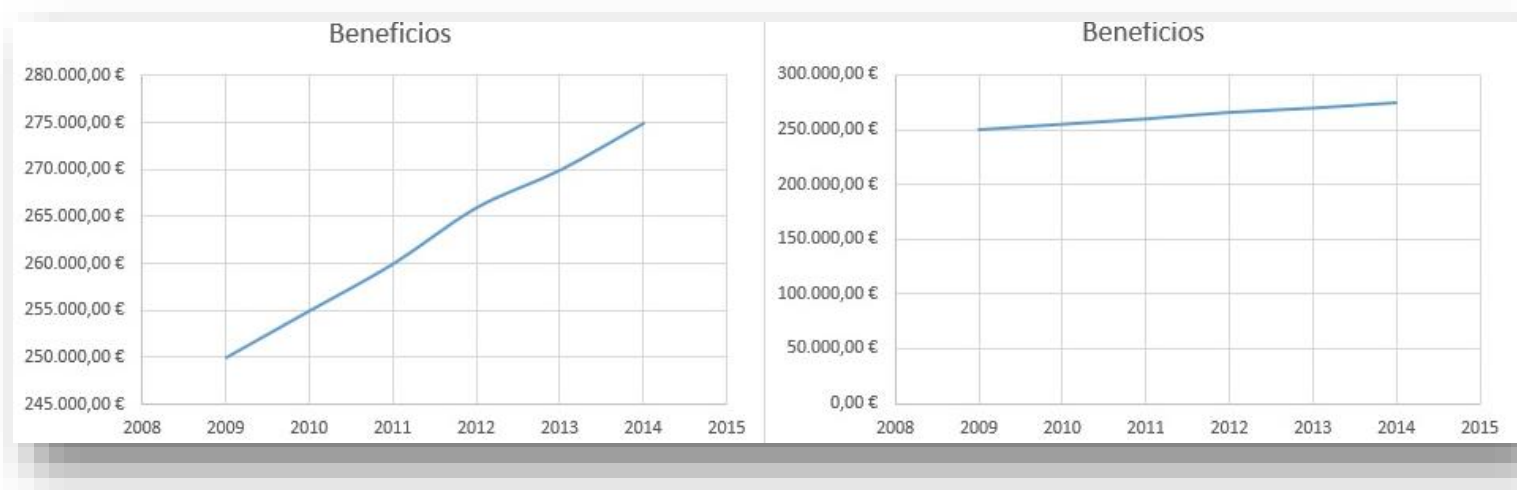


Ilustración 44: Comparativa de gráficos

4.2 Creación de Dashboard

Lo primero que tenemos que hacer es entrar en Pentaho como administrador (Ilustración 45). Si no nos hemos preocupado de crear nuevos roles de usuario, recordad que las credenciales de acceso como administrador por defecto son: User name: “admin”, Password: “password”.



Ilustración 45: Pantalla inicial Pentaho BI Server

Una vez registrado como administrador, creamos un nuevo tablero CDE, como se ve en la Ilustración 46. Esto nos mostrará el editor de Dashboard de Pentaho, el cual se divide en tres partes (Layout Structure, Datasources y Components).

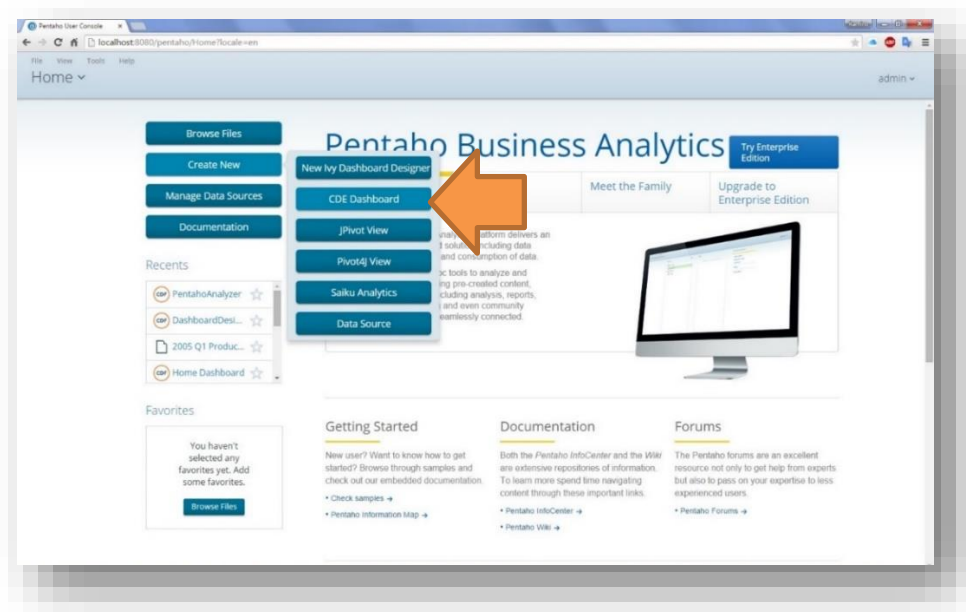


Ilustración 46: Create New, CDE Dashboard

4.2.1 Layout Structure

En este editor es donde vamos a crear nuestro Dashboard. En primer lugar definiremos la disposición visual de los componentes de nuestro tablero. Añadiremos filas o columnas, según nos convenga. En la ilustración 47 se muestra una disposición recomendada.

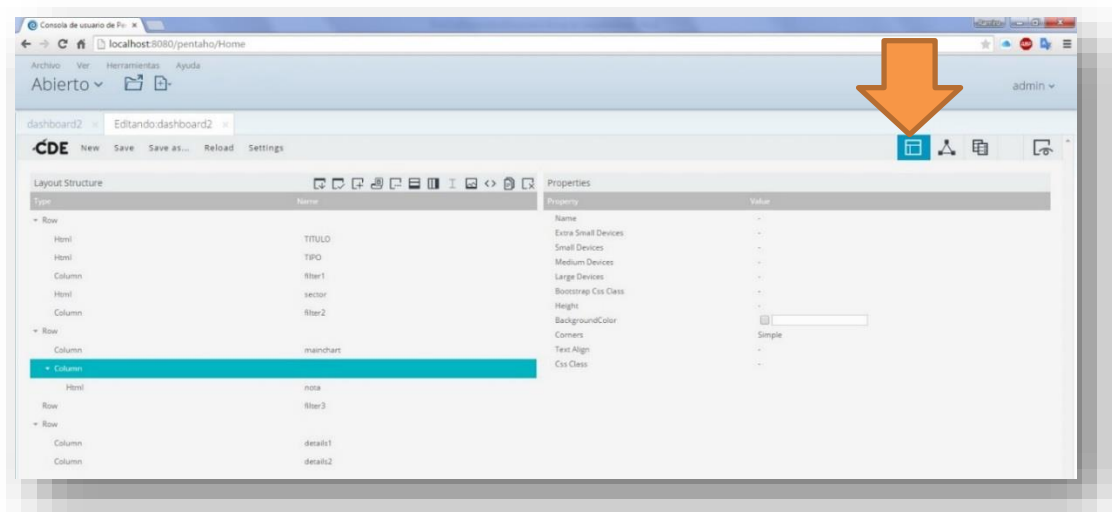


Ilustración 47: Layout Structure

4.2.2 Datasources

En la pestaña Datasources crearemos los selectores y las consultas necesarias para luego construir los gráficos, Ilustración 48.

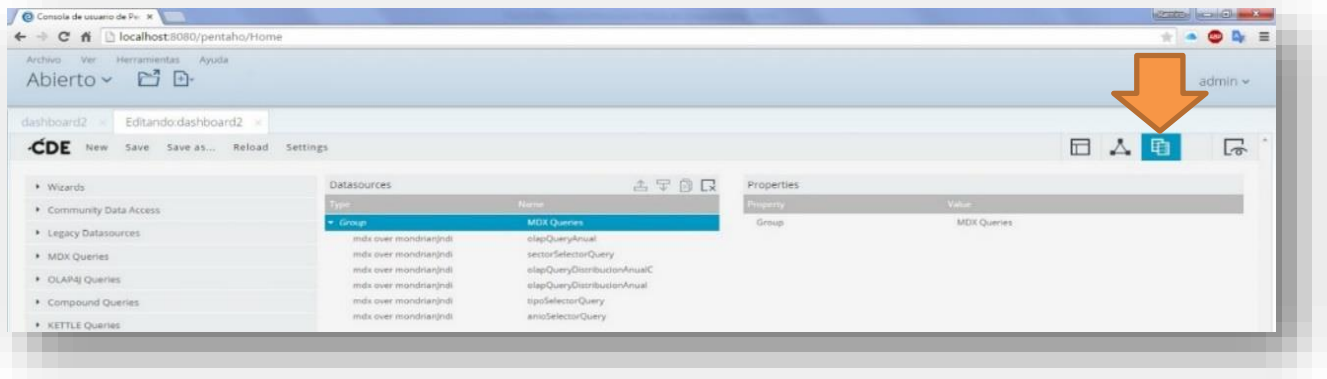


Ilustración 48: Datasources

Para crear un selector pulsaremos en la parte izquierda sobre OLAP Selector Wizard, tal y como se muestra en la Ilustración 49.

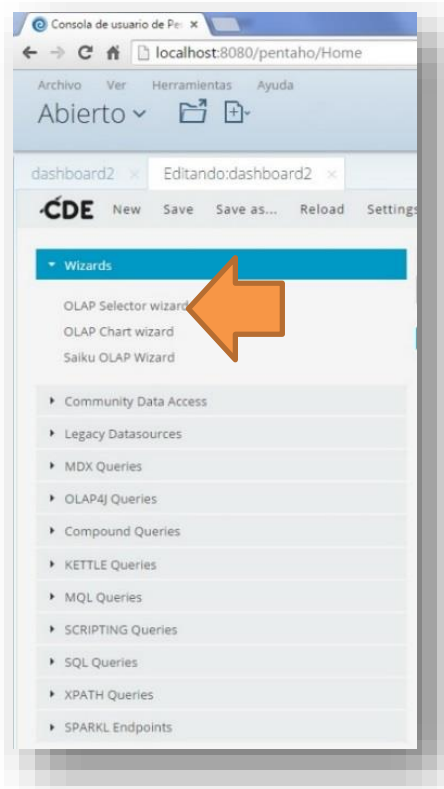


Ilustración 49: OLAP Selector Wizard 1

Posteriormente nos aparecerá la pantalla de creación de selectores, tal y como podemos apreciar en la Ilustración 50, en la que debemos seleccionar el cubo sobre el que lo queremos realizar, así como la dimensión que deseemos. Para ello, debemos arrastrar la dimensión al espacio Rows tal y como marca la flecha de la Ilustración 50.

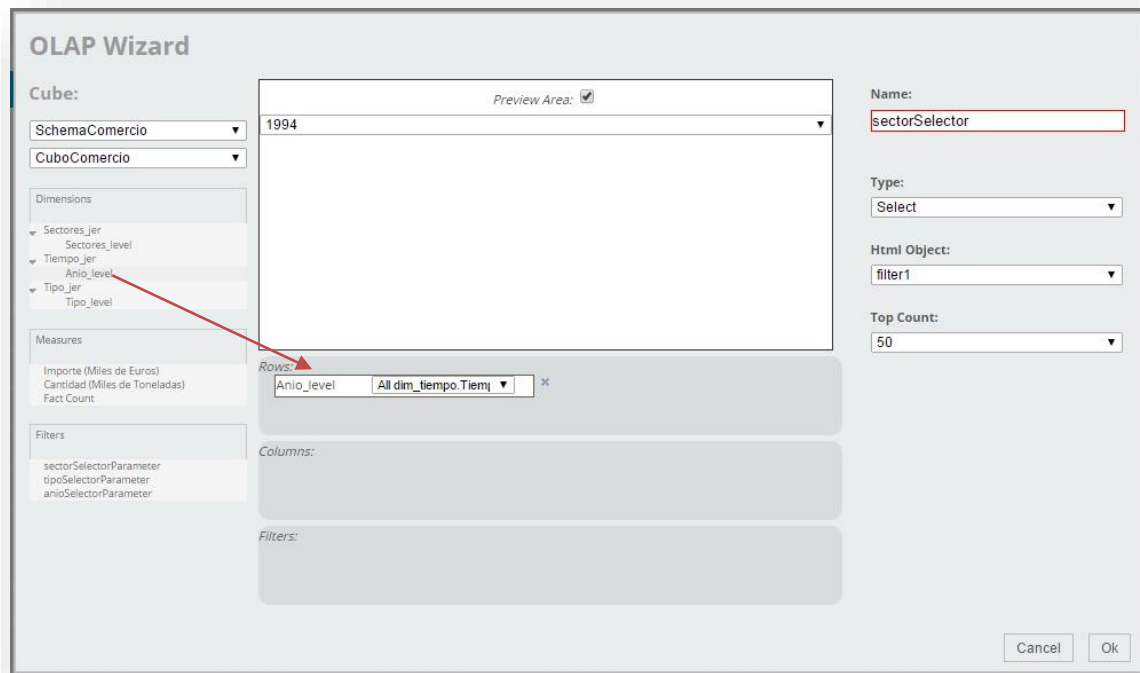


Ilustración 50: OLAP Selector Wizard 2

Seguido, en la pestaña Datasource, como se ve en la ilustración 51 realizaremos la consulta que luego nos servirá para construir los gráficos del Dashboard. Para ello utilizaremos la herramienta Saiku Analytics 2.6 (Utilizaremos esta versión puesto que no contiene errores y se pueden realizar todas las tareas que necesitamos correctamente), como también podemos ver en la Ilustración 52 (Datasource para gráfico de evolución anual).

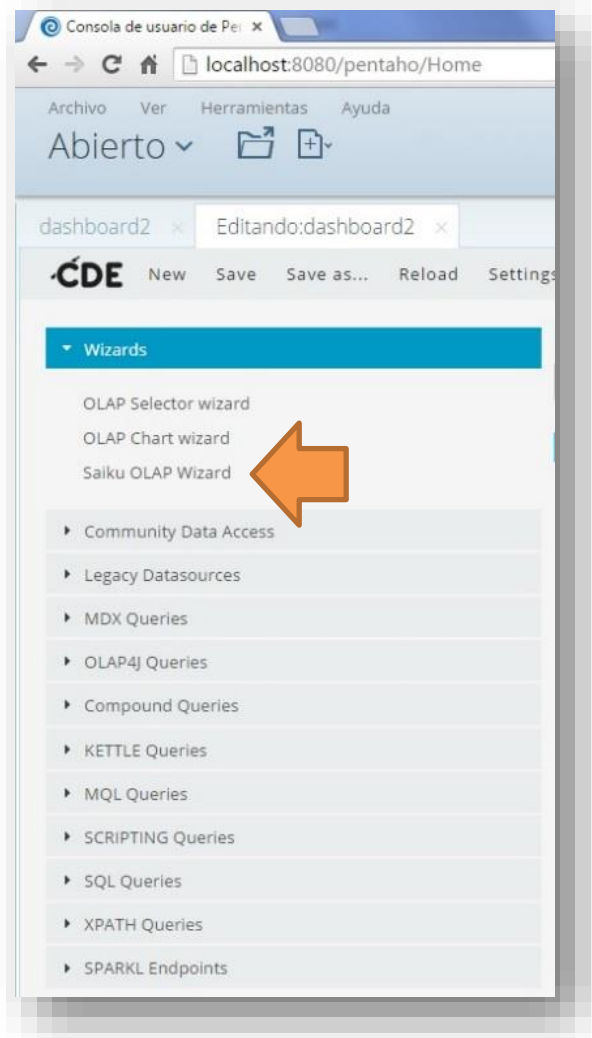


Ilustración 51: Saiku OLAP Wizard 1

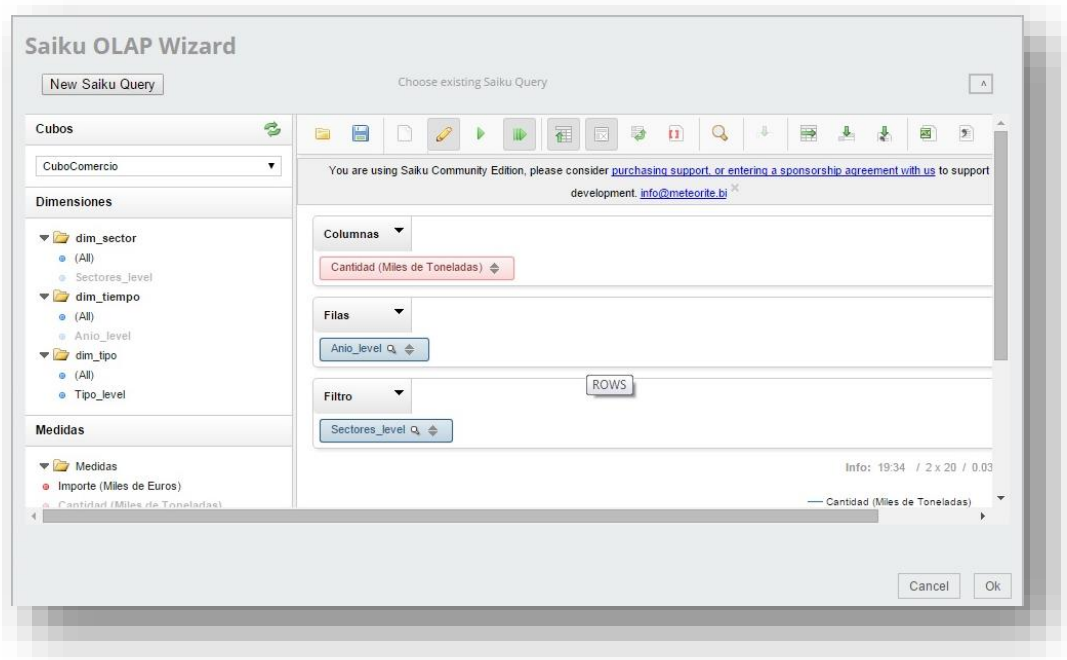


Ilustración 52: Saiku OLAP Wizard 2

Cabe destacar que en la dimensión que utilicemos como filtro debemos seleccionar una concreta para que se pueda realizar la consulta en base a ese dato específico. En nuestro caso seleccionaremos el sector “Total”, tal y como se aprecia en la Ilustración 53. Esto nos servirá para que cuando editemos la consulta manualmente sustituyamos el valor “Total” por el valor del sector correspondiente marcado en el selector sector.

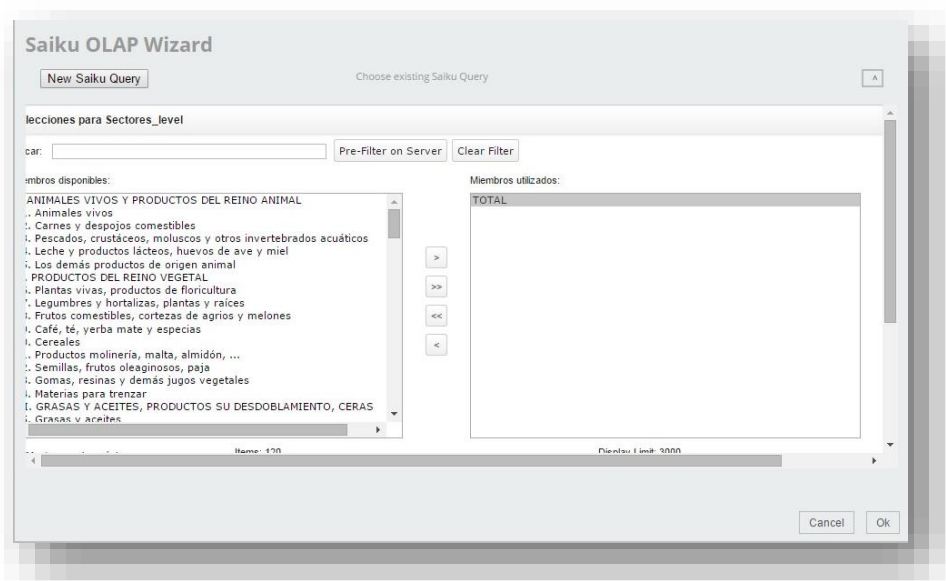


Ilustración 53: Saiku OLAP Wizard 3

Una vez tenemos creada la Query con Saiku debemos especificarle manualmente:

- El Jndi (Java Naming and Directory Interface): La Interfaz de Nombrado y Directorio Java permite descubrir y buscar objetos y datos a través de un nombre, Ilustración 54.
- Mondrian Schema, esquema al sobre el que hemos realizado la consulta tal y como se aprecia en la Ilustración 54.

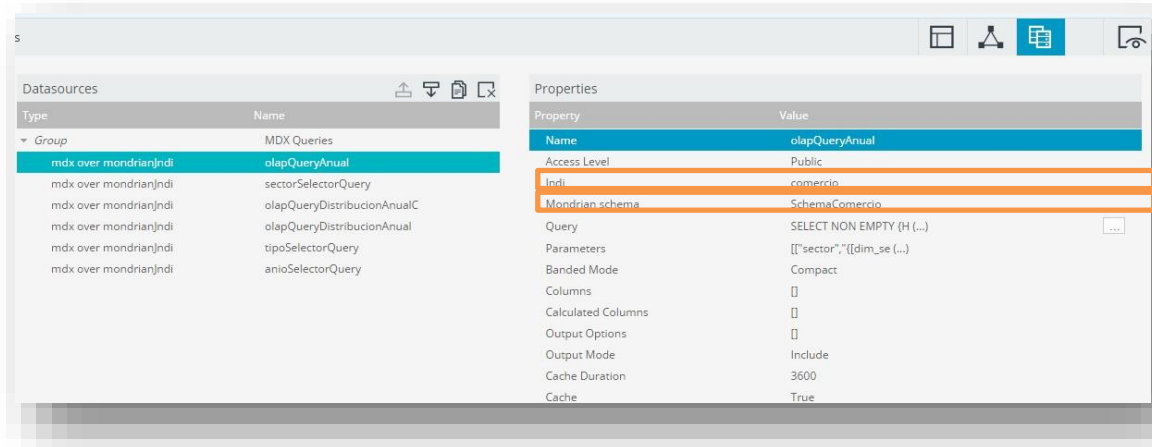


Ilustración 54: Jndi y Mondrian Schema

Para posteriormente poder enlazar los selectores con los gráficos es necesario crear los parámetros que definirán la consulta. Lo haremos modificando las dos casillas que se ven en la Ilustración 55.

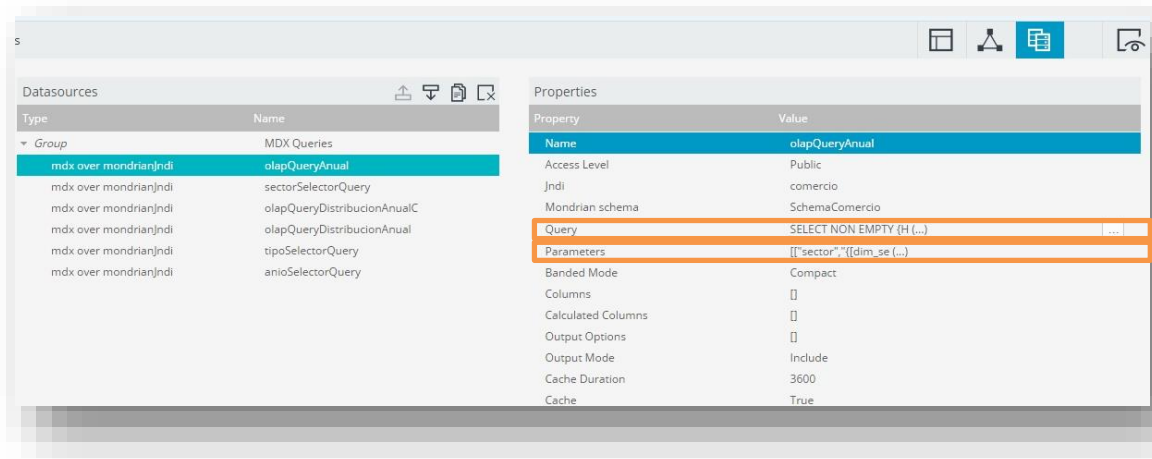


Ilustración 55: Query y Parameters

En primer lugar modificaremos la Query eliminando los parámetros estáticos que hemos seleccionado en los filtros de la consulta de Saiku (Ilustración 56) y los cambiaremos por parámetros variables que posteriormente vamos a definir (Ilustración 57).

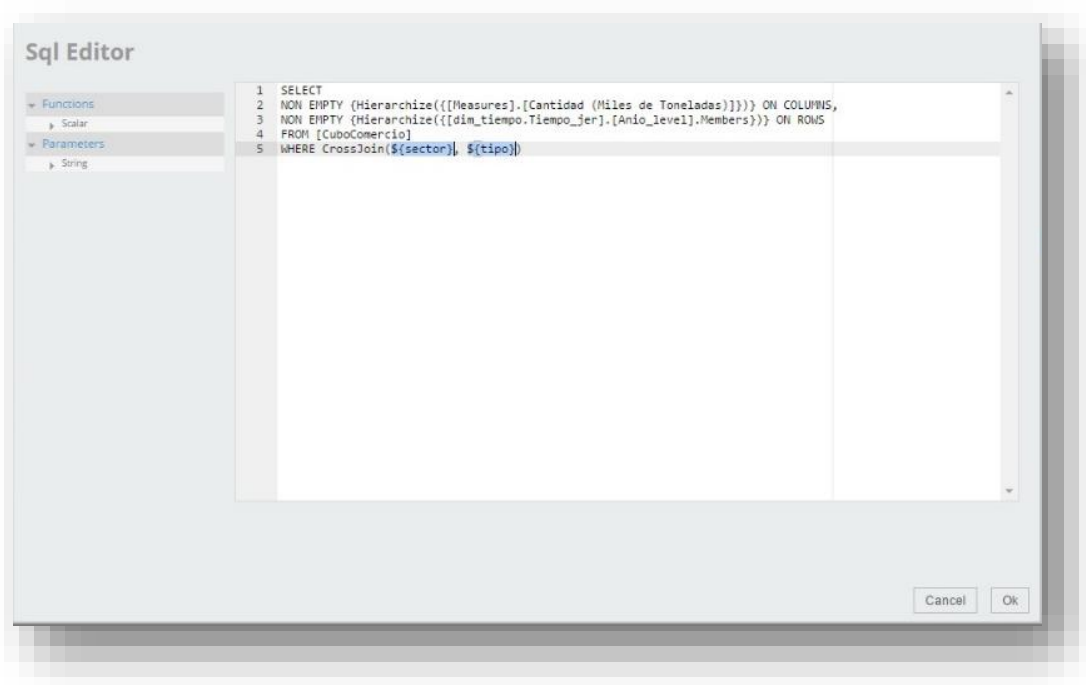


Ilustración 56: Query

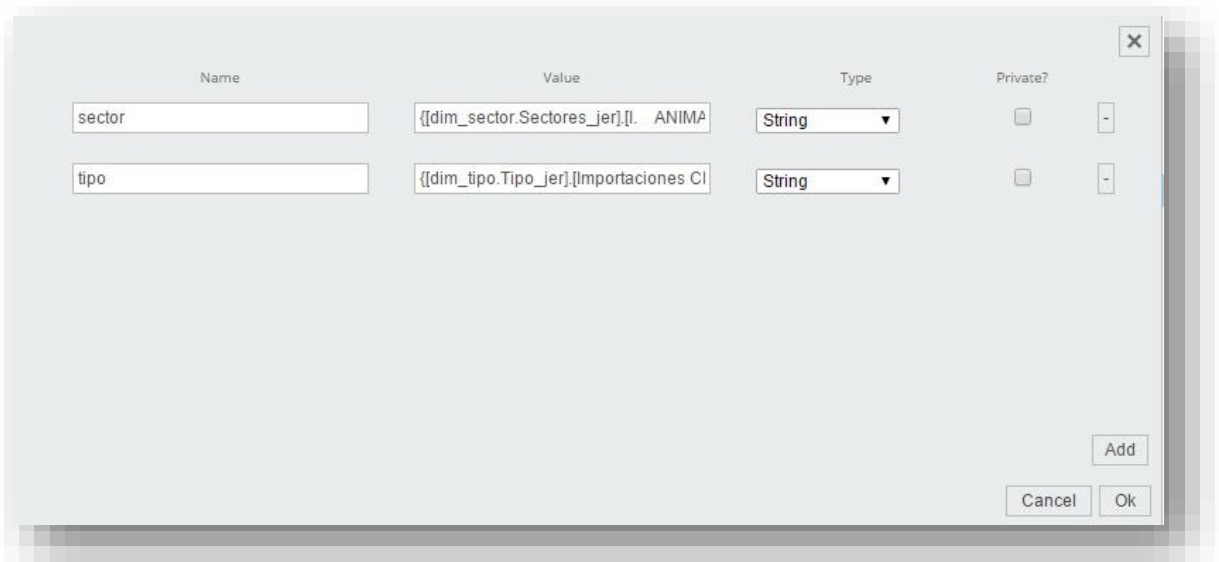


Ilustración 57: Parameters

4.2.3 Components

Una vez hecho esto solo nos queda añadir los gráficos del tipo que queramos generados con los datos obtenidos en las consultas anteriormente descritas. Tal y como se muestra en la Ilustración 58, hemos añadido dos gráficos de sector y uno lineal para mostrar los diferentes datos que pretendemos.

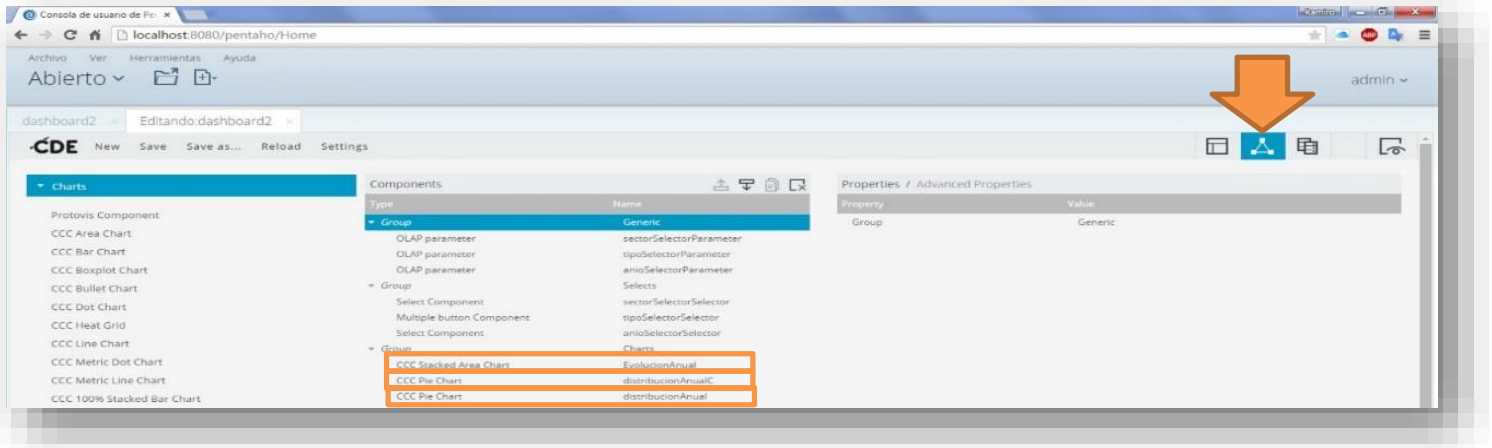


Ilustración 58: Charts

Para construir el gráfico de línea en el que mostraremos la evolución anual filtrando el tipo (Importación, Exportación) y el sector, añadimos un nuevo CCC Stacked Area Chart en el que realizaremos las siguientes modificaciones (Ilustración 59):

- Name: nombre del gráfico.
- Title: título del gráfico.
- Listeners: aquí definiremos los escuchadores del gráfico, es decir le especificaremos cuando debe actualizarse porque ha cambiado algún parámetro, en nuestro caso concreto los Listeners serán los de la Ilustración 60.
- Parameters: Especificaremos los parámetros, en concreto son los mismos que hemos añadido en la consulta creada en la parte de Datasource, Ilustración 61.
- Datasource: especificamos el origen de datos que hemos creado en el apartado Datasource correspondiente al grafico que queremos generar, en este caso es "olapQueryAnual".
- Height: asignamos una altura al gráfico.
- HTMLObject: especificamos la posición en la que queremos que se sitúe nuestro gráfico de las existentes en la parte de Layout Estructure.

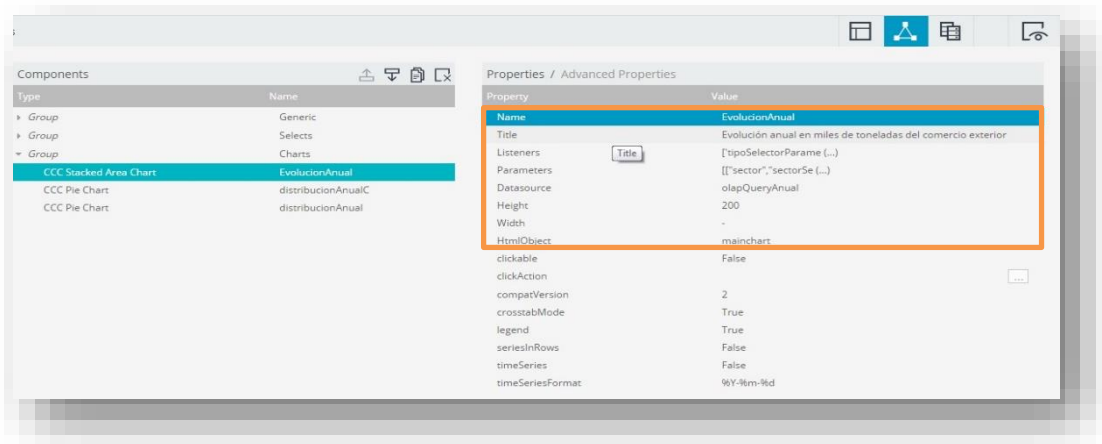


Ilustración 59: Modificaciones Chart

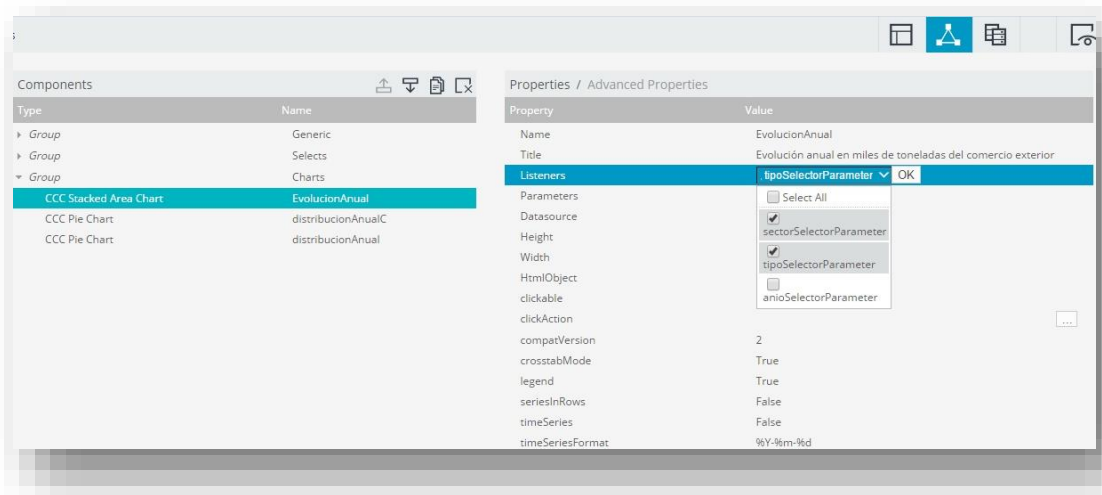


Ilustración 60: Listeners Chart

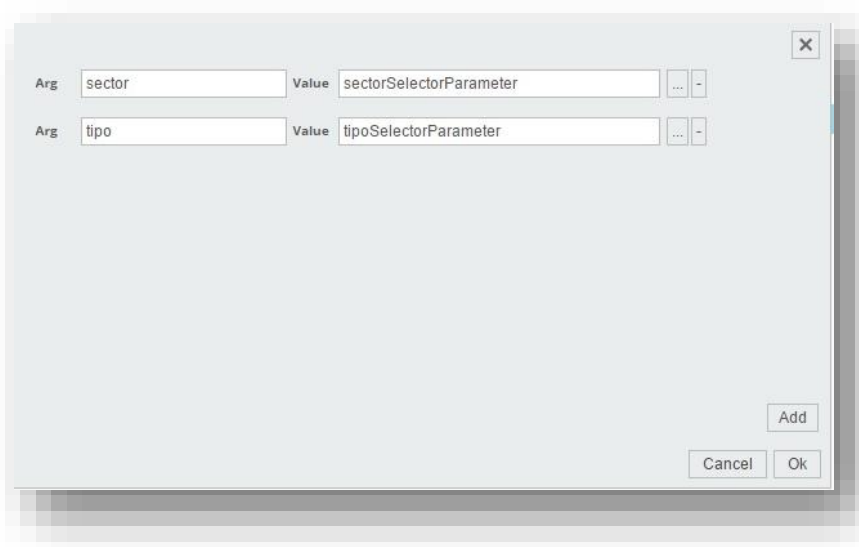


Ilustración 61: Parameters Chart

Una vez terminado el Dashboard, lo guardamos en la ruta que queramos y ya lo tenemos preparado para poder utilizarlo.

4.3 Vista del Dashboard


Mientras vamos generando las distintas partes del Dashboard podemos visualizarlo utilizando la pestaña  situada en la parte superior izquierda del editor (Ilustración 62).



Ilustración 62: Vista rápida del Dashboard

4.4 Utilización del Dashboard

Para la utilización del gráfico, el gerente de la empresa solo deberá seleccionar aquello que quiera consultar y los gráficos se actualizarán en tiempo real.

Ejemplo: Se quiere hacer un estudio de mercado sobre si sería viable establecer un negocio de venta de raíles para construir vías de trenes, este producto entraría dentro del sector: XVII. MATERIAL DE TRANSPORTE. Por tanto analizamos dicho sector para ver si hay mercado de exportación (Ilustración 63).

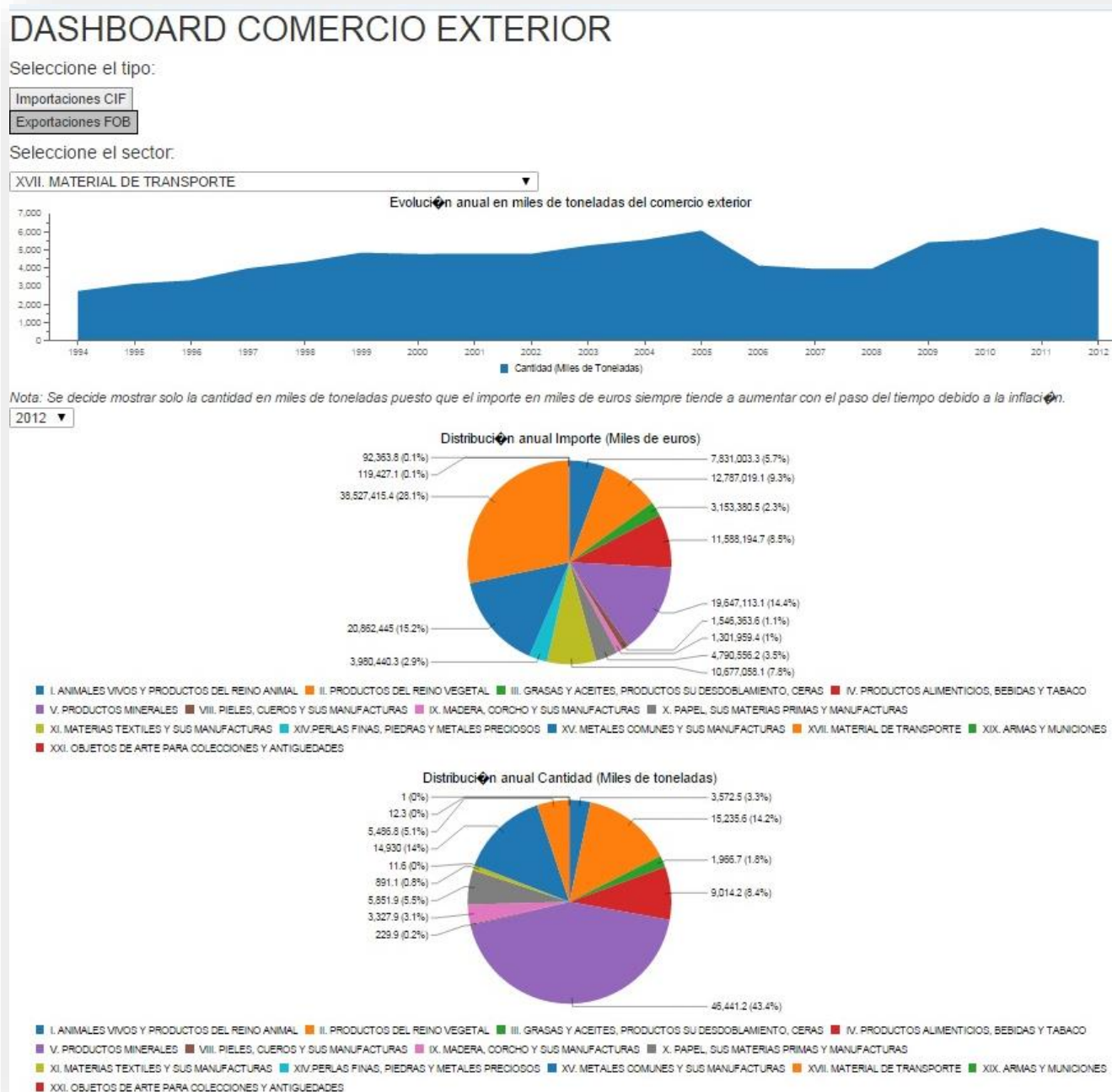


Ilustración 63: Utilización del Dashboard

Como se ve en el gráfico de evolución anual, la exportación de material de transporte ha ido creciendo de una manera más o menos constante. Se aprecia un ligero descenso alrededor del periodo de recesión española pero parece que durante los últimos tres años ha recuperado su tendencia alcista. Además, también apreciamos que en el año más reciente del que se dispone de datos (2012) el 28,1% de las exportaciones totales anuales de España son de este sector.

Por ello, todo parece indicar que hay bastante mercado en este ámbito como para establecer un negocio de venta de raíles para vías del tren.

5 Elaboración de Reportes.

5.1 Realización del reporte

Para realizar los reportes vamos a utilizar la herramienta Report Wizard (Ilustración 64). Esta herramienta nos guiará a través de un asistente con el cual podremos construir el reporte. Posteriormente pormenorizaremos un poco modificando los campos que más nos convengan de manera manual.

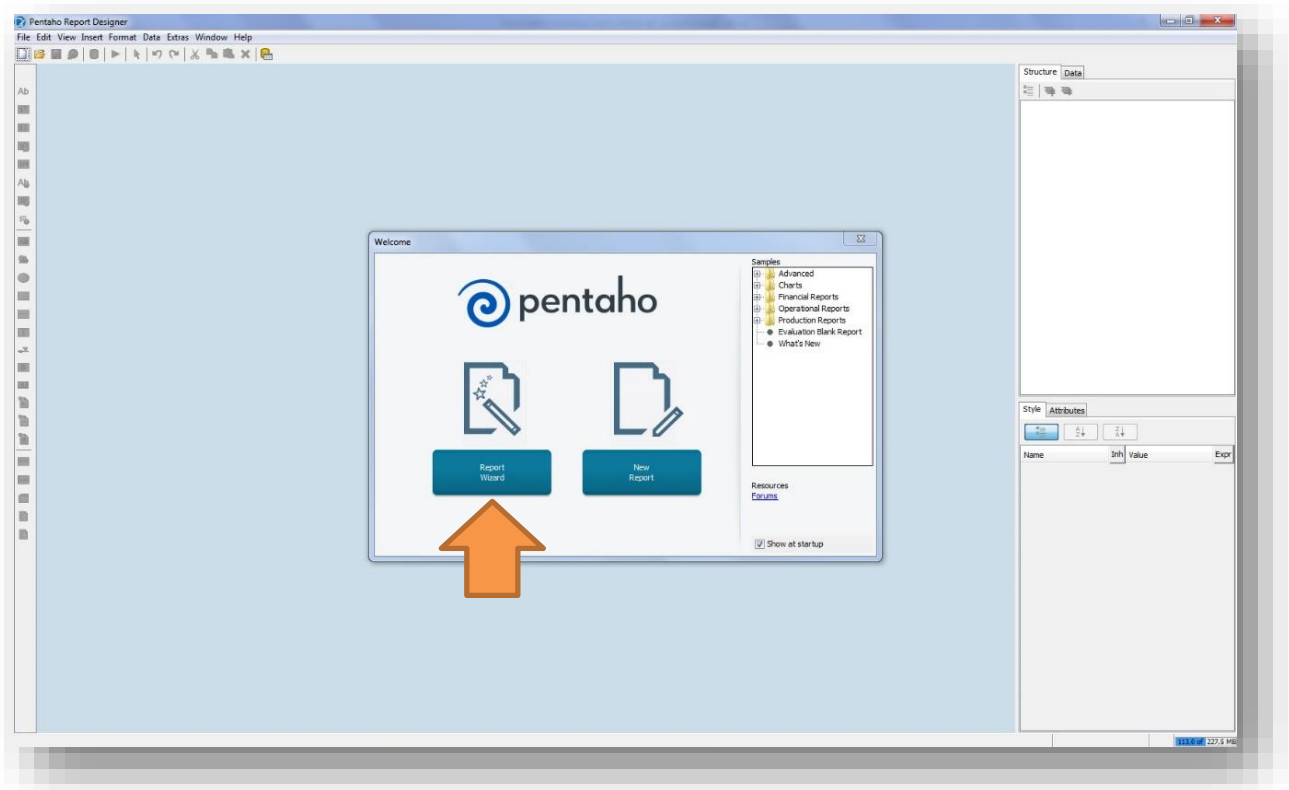


Ilustración 64: Pantalla inicial Pentaho Report Designer

5.1.1 Look and Feel

En segundo lugar seleccionamos una plantilla que defina el diseño de nuestro reporte (Ilustración 65).

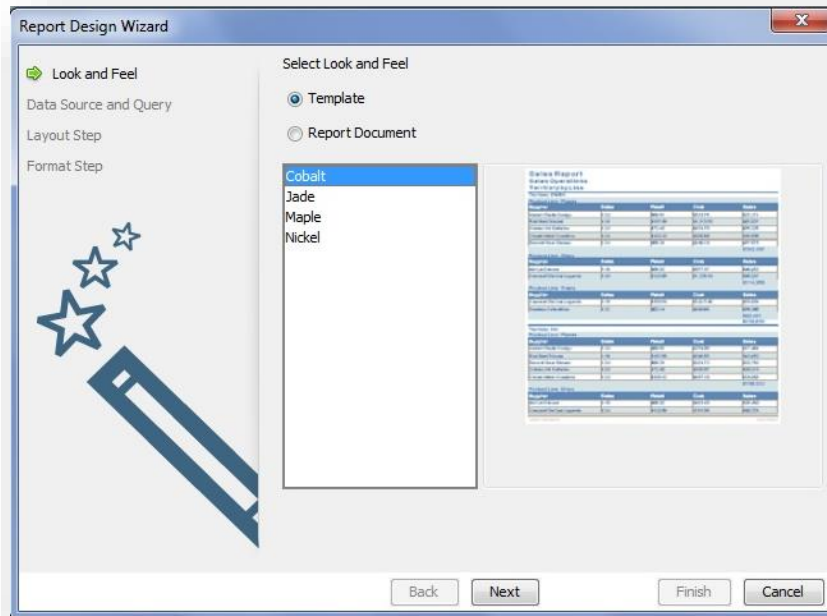


Ilustración 65: Report Design Wizard Look and Feel

5.1.2 Data Source and Query

Añadimos un nuevo Datasource, tal y como se muestra en la Ilustración 66.

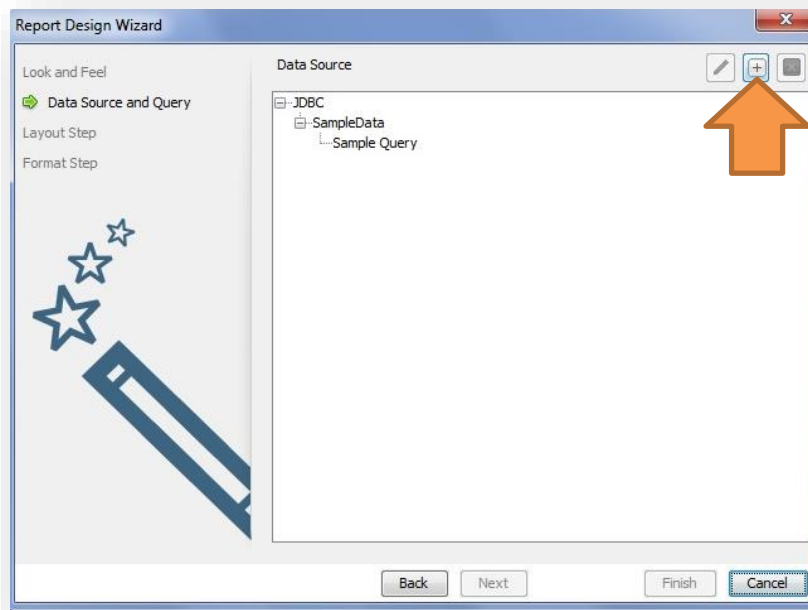


Ilustración 66: Report Design Wizard Data Source and Query 1

Seleccionamos JDBC en el menú desplegable que muestra la Ilustración 67.

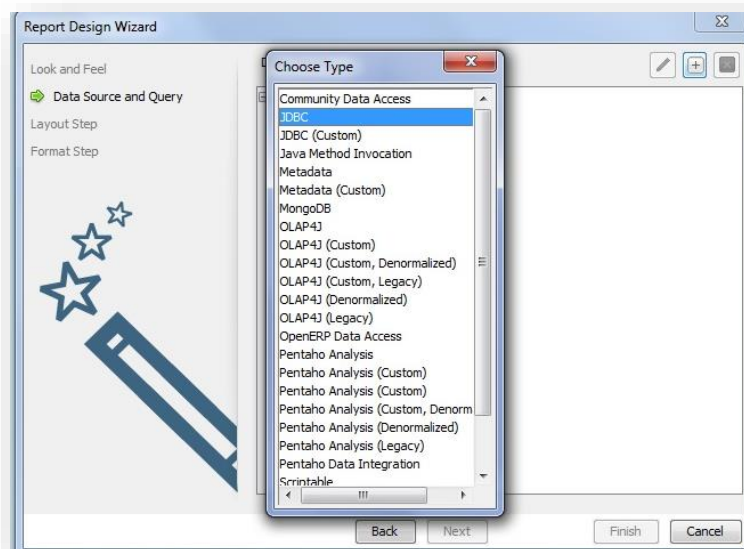


Ilustración 67: Report Design Wizard Data Source and Query 2

Posteriormente rellenamos las credenciales de acceso a nuestra base de datos MySQL tal y como se indica en la Ilustración 68.

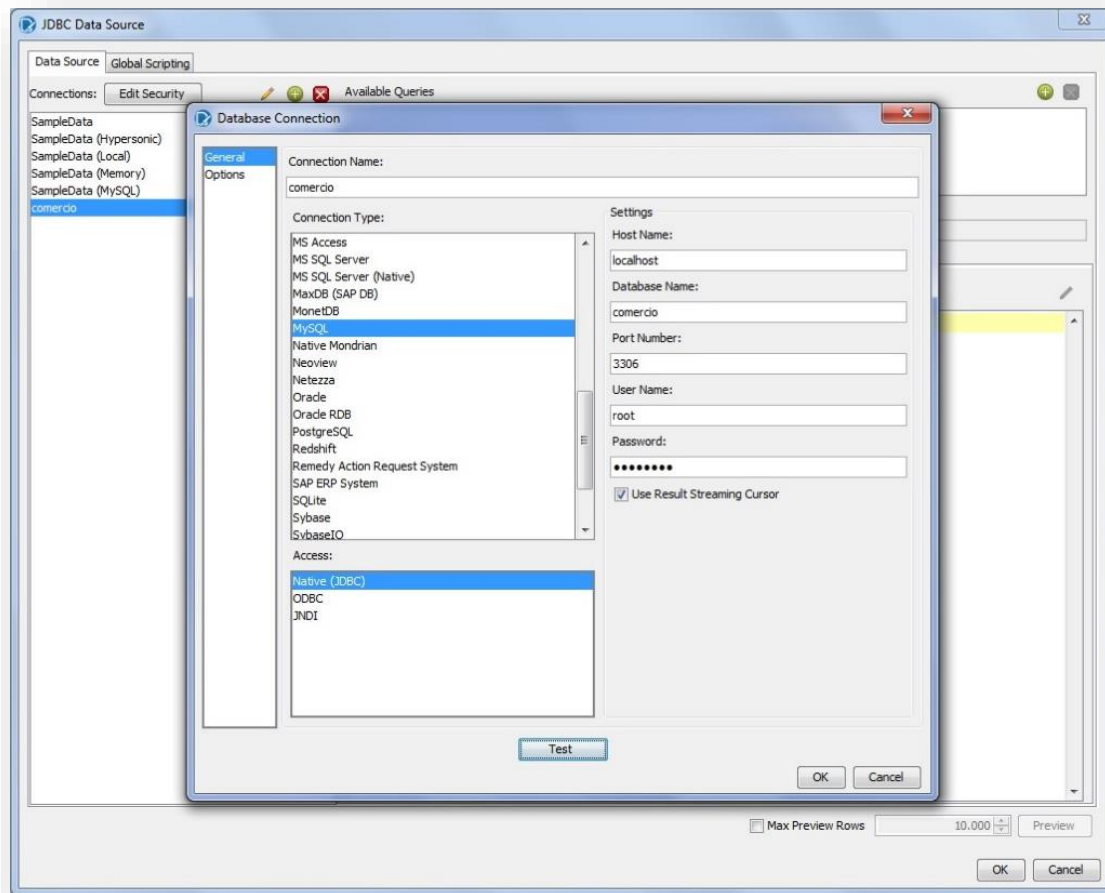


Ilustración 68: Query Database Connection

Comprobamos que la conexión está correctamente establecida haciendo un test de la misma y comprobando que éste es satisfactorio (ver Ilustración 69).

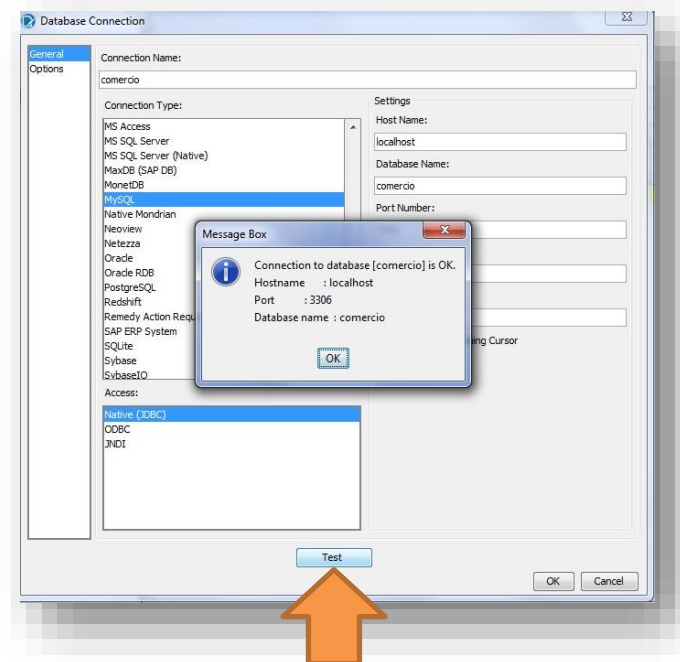


Ilustración 69: Query Database Connection Test

Una vez tenemos creada la conexión, tal y como se aprecia en la Ilustración 70, añadimos una nueva Query con el nombre que más nos convenga.

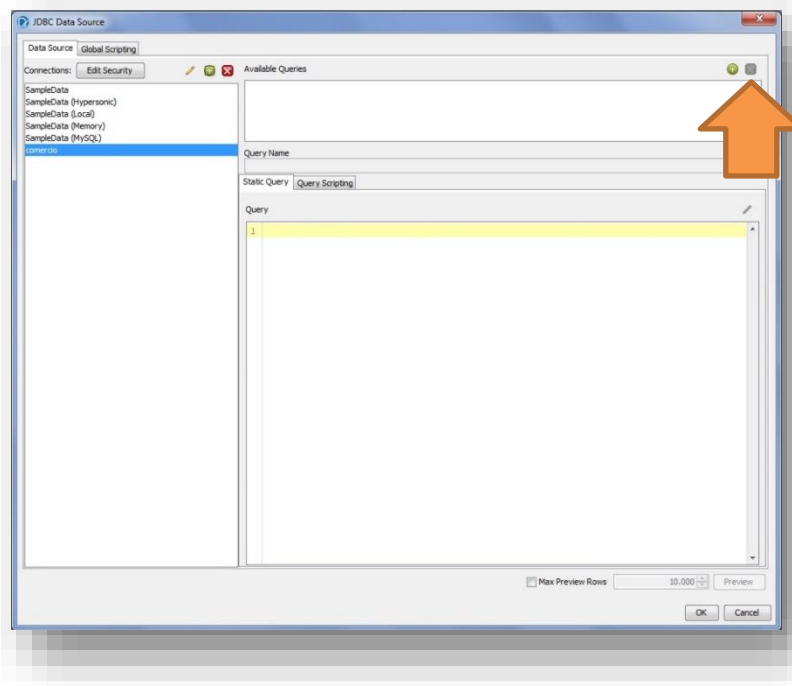


Ilustración 70: New Query 1

Una vez creada la nueva Query, tenemos dos opciones, o la escribimos directamente en SQL, o utilizamos el editor (Ilustración 71) para construir nuestra consulta de una forma un poco más intuitiva.

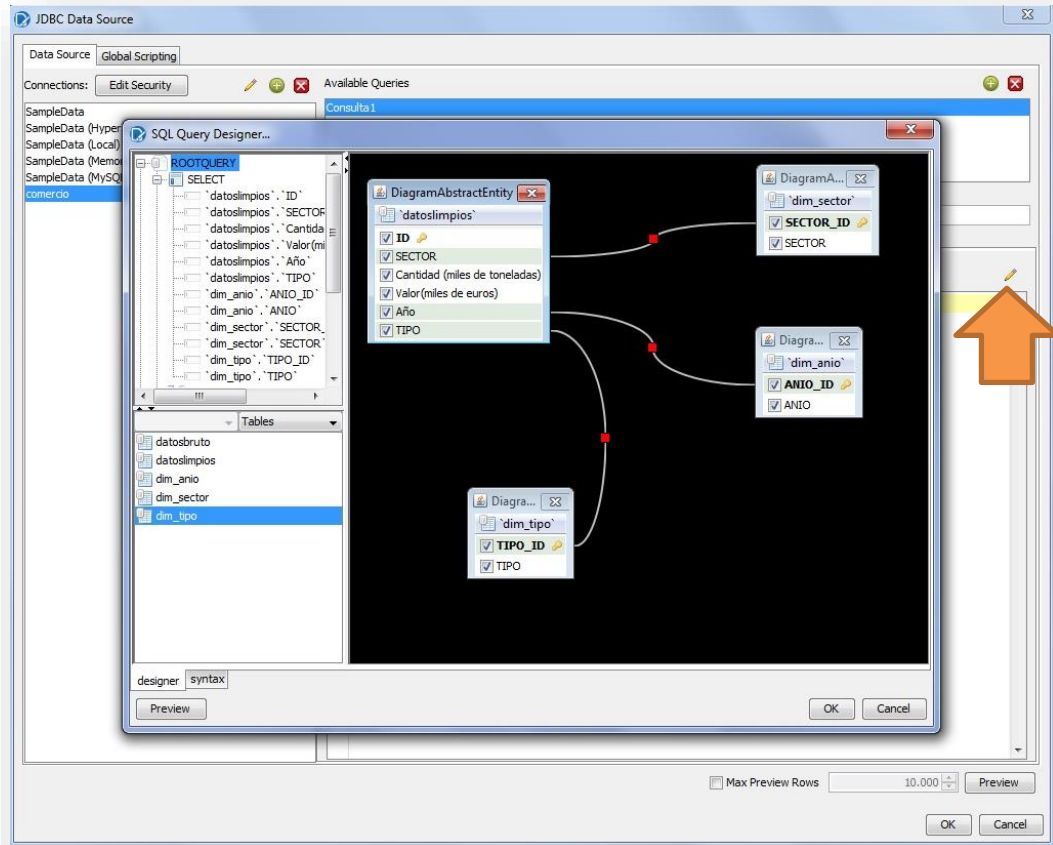


Ilustración 71: SQL Query Designer

Una vez estamos satisfechos con la consulta construida a través del editor, haremos clic en el botón “OK” y pasaremos a la siguiente pantalla en la que se ve la consulta generada (Ilustración 72).

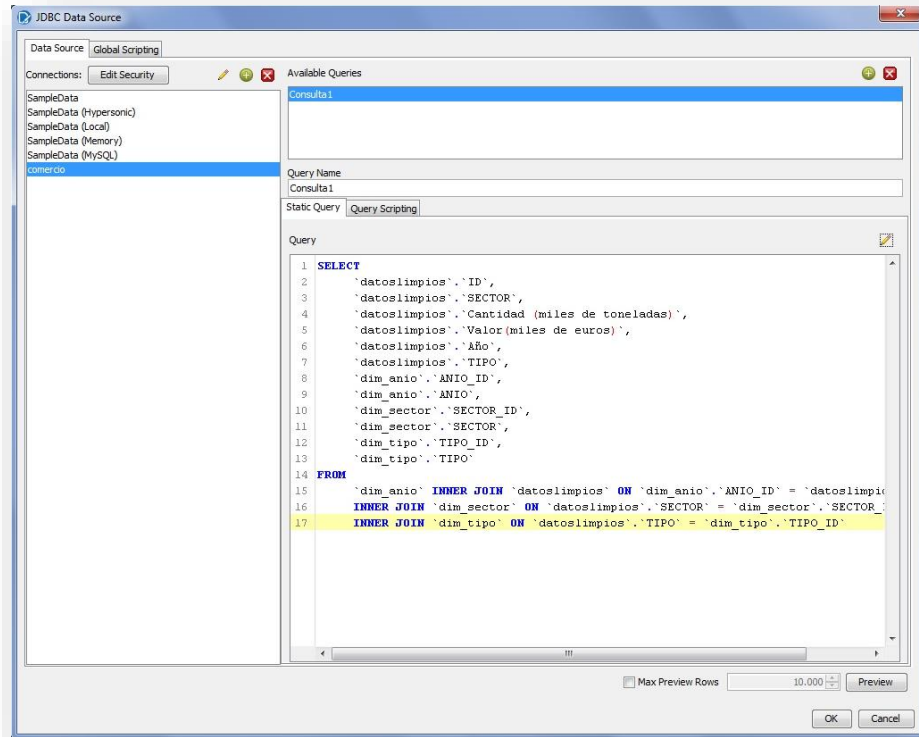


Ilustración 72: New Query 2

Continuaremos haciendo clic en “OK” y pasaremos a la ventana del asistente en la que nos aparecerá seleccionada la consulta que hemos realizado en el Datasource añadido (Ilustración 73).

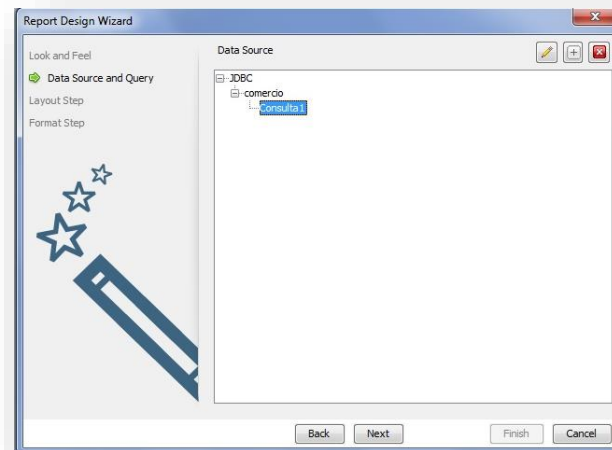


Ilustración 73: Report Design Wizard Data Source and Query 3

5.1.3 Layout Step

En el siguiente paso vamos a diseñar nuestro reporte. Es aquí donde definiremos los argumentos seleccionados y aquellos argumentos por los que queremos agrupar la selección. En este caso seleccionamos cantidad, valor y tipo, y agrupamos los datos en torno a año y sector, tal y como se aprecia en la Ilustración 74.

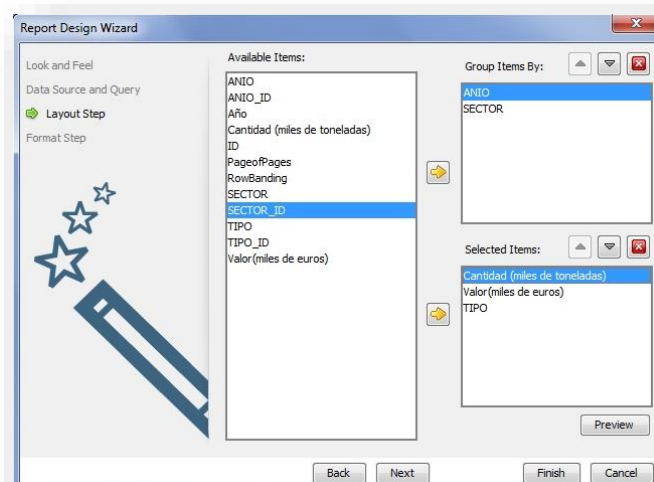


Ilustración 74: Report Design Wizard Layout Step

Por último, finalizamos el asistente y se nos mostrará el reporte en el que podemos modificar los campos manualmente para personalizarlo y ajustarlo a nuestras necesidades. En nuestro caso, añadiendo el título y cambiando el formato de salida de los datos, nos queda el reporte de la Ilustración 75.

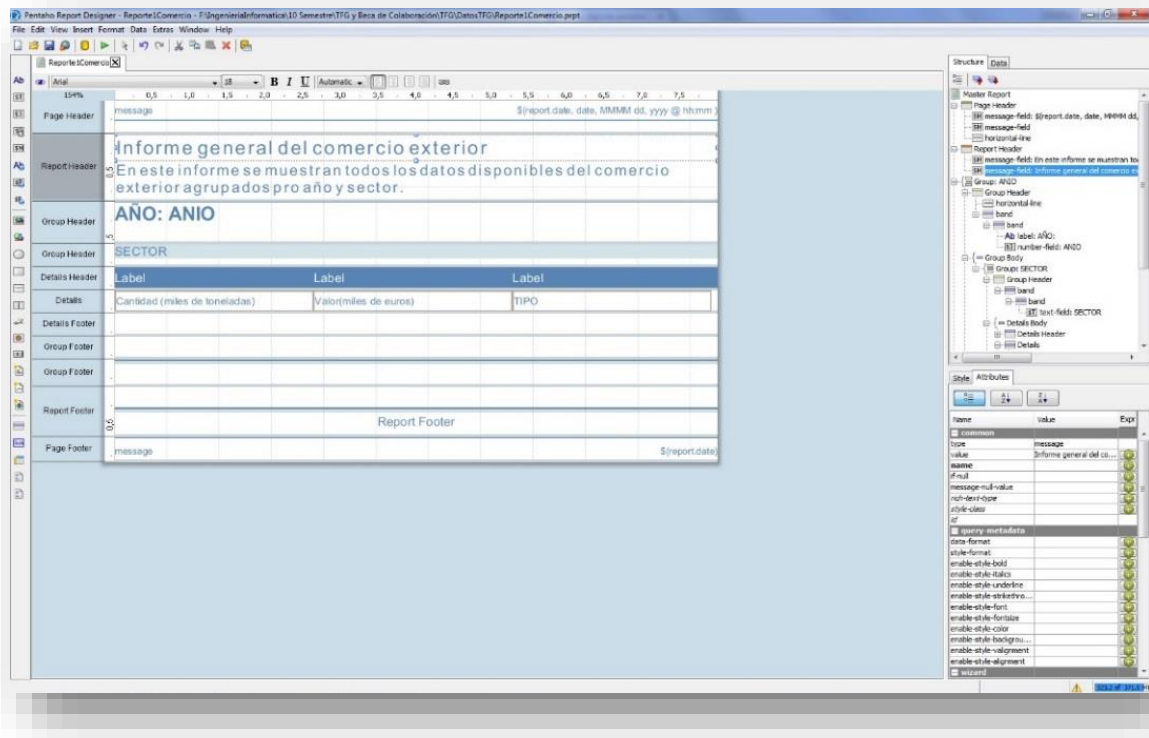



Ilustración 75: Vista general Reporte

Una vez tenemos conformado el reporte, haciendo clic en  lo ejecutaremos en los distintos formatos posibles de salida (Ilustración 76), como por ejemplo puede ser un PDF. En la Ilustración 77 se muestra un Print Preview de nuestro reporte.

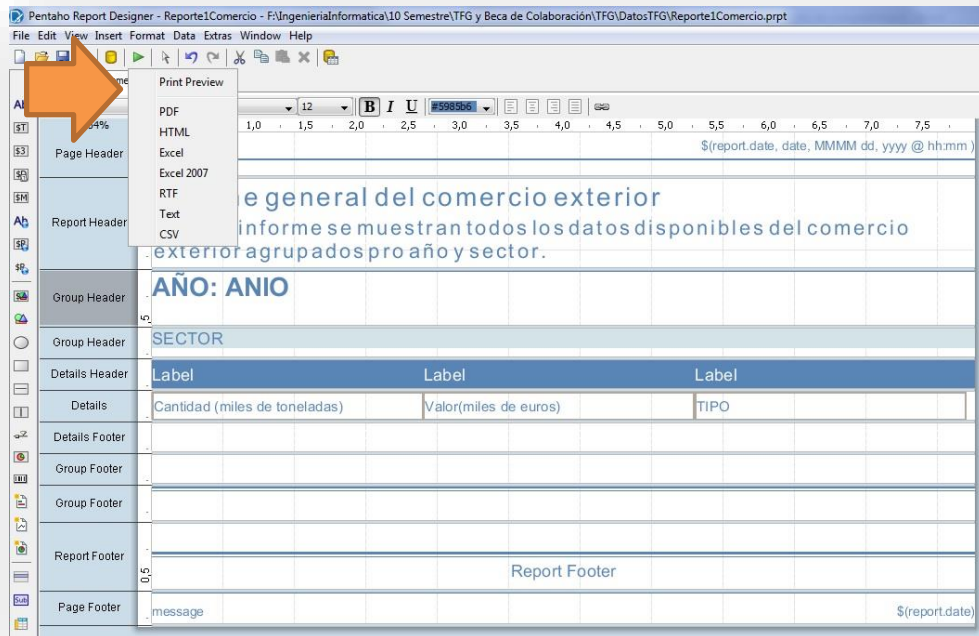


Ilustración 76: Run Report

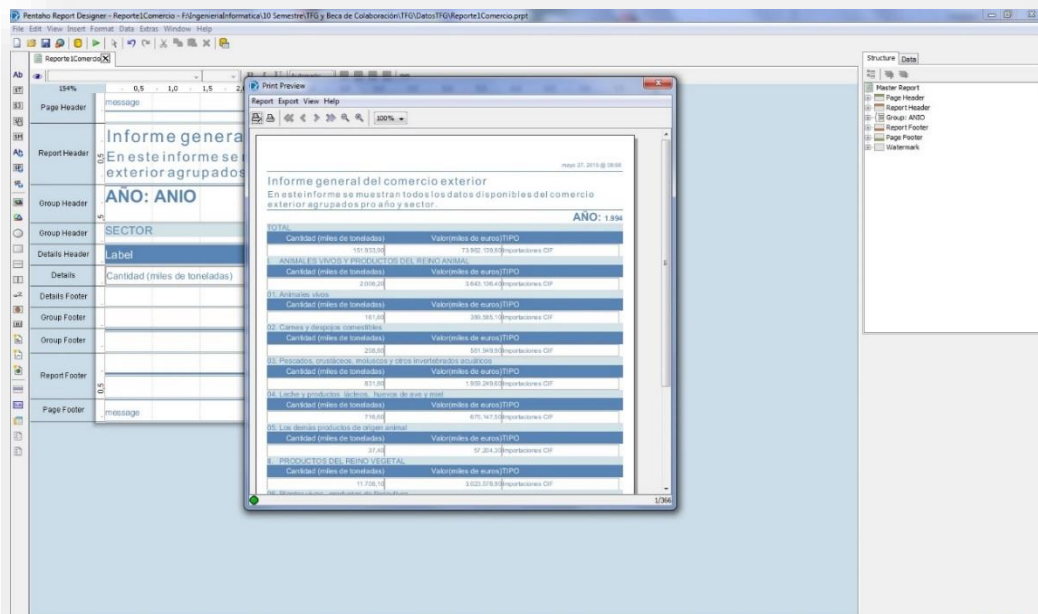


Ilustración 77: Print Preview Report

5.2 Conclusión

Los reportes, aunque no son una herramienta que nos permita tener un dinamismo propio del análisis de datos, como puede ser el caso del Dashboard, es necesario generarlos puesto que es la forma de hacer fehaciente la toma de una decisión. Son documentos que se emplean para respaldar las decisiones tomadas ante aquellos interlocutores a los que se deba rendir cuentas.

6 Conclusiones y líneas futuras

Business Intelligence es un campo que últimamente ha tenido mucho auge en la mayoría de las empresas que manejan gran cantidad de datos estructurados. Esto es debido a que es evidente que una herramienta con estas características es capaz de aportar suficiente valor añadido a las compañías como para que las mismas decidan invertir en este ámbito cada vez más parte de sus beneficios.

Se podría concluir que una empresa que dispone de un software de análisis de datos de este tipo, perfectamente integrado y con una buena interpretación de la gerencia, adquiere una ventaja competitiva con respecto a los que carecen de él.

En mi opinión, el futuro nos depara una nueva dimensión de actuación de todo este tipo de herramientas, puesto que para el desempeño de una compañía no solo son importantes los datos internos de la misma (stock, ventas, producción, salarios,...) sino que también son relevantes los datos externos, como por ejemplo pueden ser las opiniones de los clientes en las redes sociales, o cualquier otro tipo de datos relacionados con la misma. Este tipo de datos es complicado agruparlos puesto que son datos desestructurados pero bajo mi punto de vista es la dirección y el sentido hacia el que navega toda esta tecnología.

Creo que este tipo de herramientas no se puede quedar en el puro análisis de datos, sino que también es presumible que se exigirá cada vez con más frecuencia en el mundo empresarial que este tipo herramientas sea capaz de prever el futuro. Prever el impacto que tendrá tomar una decisión u otra en la propia organización para así enderezar el rumbo en un sentido u otro.

En el caso concreto de Pentaho CE he de decir que me ha resultado fácil e intuitivo realizar gran parte del proyecto y que en lo que quizá haya sido más complicado profundizar sea en la parte de Data Integration puesto que las posibilidades de origen de datos y las herramientas que ofrece para su tratamiento son inmensas. Es una herramienta que al principio cuesta utilizar sobre todo si no estás familiarizado con este tipo de software pero he de decir que el tiempo de aprendizaje es bastante razonable.

Para terminar haré una reflexión sobre éste mi Trabajo Fin de Grado:

Estaré satisfecho, si en un futuro vuelvo a leer estas hojas y descubro que podría haberlo hecho mucho mejor, señal de que habré evolucionado como persona y de que mis conocimientos también lo habrán hecho en consonancia. Si algo he podido aprender en mi corta vida es que jamás hay que perder la ilusión por aprender, sea cual sea la disciplina.

7 Referencias

- [1] «INE» [En línea]. Available: <http://www.ine.es/>. [Último acceso: Junio 2015].
- [2] «Pentaho Community Edition» [En línea]. Available: <http://community.pentaho.com/>. [Último acceso: Junio 2015].
- [3] «RapidMiner» [En línea]. Available: <https://rapidminer.com/>. [Último acceso: Junio 2015].
- [4] «SpagoBI» [En línea]. Available: <http://www.spagobi.org/>. [Último acceso: Junio 2015].
- [5] «Eclipse BIRT Project» [En línea]. Available: <http://www.eclipse.org/birt/>. [Último acceso: Junio 2015].
- [6] «OpenI» [En línea]. Available: <http://openi.org/>. [Último acceso: Junio 2015].
- [7] «Palo» [En línea]. Available: <http://sourceforge.net/projects/palo/>. [Último acceso: Junio 2015].
- [8] «QLR Manager» [En línea]. Available: <http://www.qlrmanager.com/>. [Último acceso: Junio 2015].
- [9] «Oracle BI» [En línea]. Available: <http://www.oracle.com/us/solutions/business-analytics/business-intelligence/overview/index.html>. [Último acceso: Junio 2015].
- [10] «Microsoft SQL Server BI Edition» [En línea]. Available: <http://www.microsoft.com/es-es/server-cloud/products/sql-server-editions/sql-server-business-intelligence.aspx>. [Último acceso: Junio 2015].
- [11] «SAP Business Objects» [En línea]. Available: <http://www.sap.com/spain/pc/analytics/business-intelligence.html>. [Último acceso: Junio 2015].
- [12] «Panorama Software» [En línea]. Available: <http://www.panorama.com/>. [Último acceso: Junio 2015].
- [13] «Pentaho BI» [En línea]. Available: <http://www.pentaho.com/>. [Último acceso: Junio 2015].

- [14] «Documentación Pentaho Community» [En línea]. Available:
<http://community.pentaho.com/>. [Último acceso: Junio 2015].
- [15] «Instalador de Java Development Kit 7» [En línea]. Available:
<http://www.oracle.com/technetwork/es/java/javase/downloads/jdk7-downloads-1880260.html>. [Último acceso: Junio 2015].
- [16] «MySQL» [En línea]. Available: <https://www.mysql.com/>. [Último acceso: Junio 2015].
- [17] J. Conesa Caralt y C. D. Josep, Introducción al Business Intelligence, UOC.
- [18] «MySQL connector Java» [En línea]. Available:
<http://dev.mysql.com/downloads/connector/j/>. [Último acceso: Junio 2015].